Evolution of MHC diversity in cetacean species environment: implications for the role of a pathogen

Vassilakos, Demetrios

Evolution of MHC diversity in cetacean species; implications for the role of a pathogen environment

A thesis submitted to the University of Durham for the Degree of Doctor of Philosophy

2004

Demetrios Vassilakos

School of Biological and Biomedical Sciences
University of Durham
Supervisor: Dr. Rus Hoelzel

ABSTRACT

Previous studies have indicated that the evolution of diversity in the immune system genes responsible for antigen presentation in terrestrial mammals is driven by selection. However, it is not clear whether mammals in the marine environment would experience the same selective pressures as terrestrial mammals, given their different pathogen environment.

The diversity and pattern of radiation of exon-2 (peptide binding region) of the *DQB1* locus of Class II MHC molecule was investigated in a broad range of cetacean species, including an extended sample of six focal species (three *Mysticeti* and three *Odontoceti* species). The role of natural selection and of the evolutionary history of this locus was evaluated based on phylogenetic and genetic distance analyses, in order to assess the following hypotheses: i) the marine pathogen environment presents comparable selection pressures to those of the terrestrial environment, ii) social behaviour and structure will affect pathogenic pressure and iii) populations of cetacean species with a world-wide distribution across different habitats and geographic regions are under differential selective pressure.

The phylogenetic comparison of the cetacean species was consistent with the trans-species evolution pattern described for terrestrial mammals. Furthermore, high non-synonymous to synonymous substitution rates suggest that polymorphism at this locus in cetaceans is maintained by natural selection. The higher number of trans-species lineages and non-synonymous substitution rates exhibited by social species suggest that social behaviour and social structure may affect pathogenic pressure. Population differentiation according to *DQB1* locus reveals a contrasting pattern to that inferred by neutral markers (microsatellite DNA) supporting the hypothesis that habitat and geographic regions may place populations under differential selection pressure.

The results of the present investigation suggest that the pattern of evolution of the immune response in cetaceans is similar to that in terrestrial mammal species.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1 – General Introduction

Chapter 2 - Materials and methods

Chapter 3 – Evolution of MHC *DQB1* exon-2 in cetaceans

Chapter 4 – Effects of cetacean life history on *DQB1* evolution

Chapter 5 – *DQB1* variation and population genetic structure in cetaceans

Chapter 6 – General discussion

LIST OF TABLES

DECLARATION

The material contained within this thesis has not been previously submitted for a degree at the University of Durham or any other university. The research reported here has been conducted by the author unless indicated otherwise

ACKNOWLEDGMENTS

It is funny sometimes how the brain works, and I say that because after 4 years of PhD the first thing that comes to my mind when I think of Rus is a door, actually the door of his office. Every single time I had to speak to him, I had to knock the door, open it, lean against it and ask him a question about, sometimes, the most irrelevant subject that came to my mind. Then close the door and leave, trying very hard to understand what he was talking about. However, this subconscious correlation is far from being trivial because this is exactly what Rus was doing, as every-time he would open a different small door for me and let me explore; that is just how this PhD was accomplished. Rus was always there for me, whatever the reason and whatever the question. If I had the chance to re-do my PhD, start all the way from the beginning with all the money in the world to do whatever my perverted biological ambition could think, well, I would do exactly the same PhD, not only due to the actual project involved but due to my supervisor, A. R. Hoelzel. Rus, thank you for all your help and advice.

I strongly believe that in the front page of this thesis I should also include the names of Collin Nicholson, Anna Fabiani, Stefania Gaspari, Ana Topf, Ada Natoli, Courtney Nicholson, Carlos de Luna, Fiona Hatchel, Dan Engelhaupt, Courtney Nicholson, Laura Corrigan, Andy Foote, Despoina Kaloriti, Dimitris Kolotouros, Foivos Pikros, Dimitris Xydadis, Vasilis Pagmantidis, Diletta Castelani, Irini Papakammenou, Nikolaos Galiatsatos, Kostantinos Kyritsis, Dimitrios Xenias and Nektaria Pouli. The friendship of these people is very valuable to me and I will never forget them.

I am grateful to NERC for funding this project and also to the people of Ustinov College who helped me a lot during my last year. I would also like to thank all the people in the Department of Biological and Biomedical sciences for helping me throughout the period of this PhD and making me feel very welcomed from the very first day I arrived in Durham.

I would also like to thank Georgos Vassilakos and Maro Vassilakou for believing in me, Nikola Vassilako, you are my inspiration in everything I do. I dedicate this work to my family.

Marina Kolokonte, thank you for your proof reading and thank you for showing me how beautiful life can be.

Last but not least, I would like to thank Professor John Trowsdale and Dr. Paul Hunt for their valuable corrections and advice.

"May there be many beds of blue chrysanthemums
blooming in our gardens...!"
J. Klein

CHAPTER 1:

GENERAL INTRODUCTION

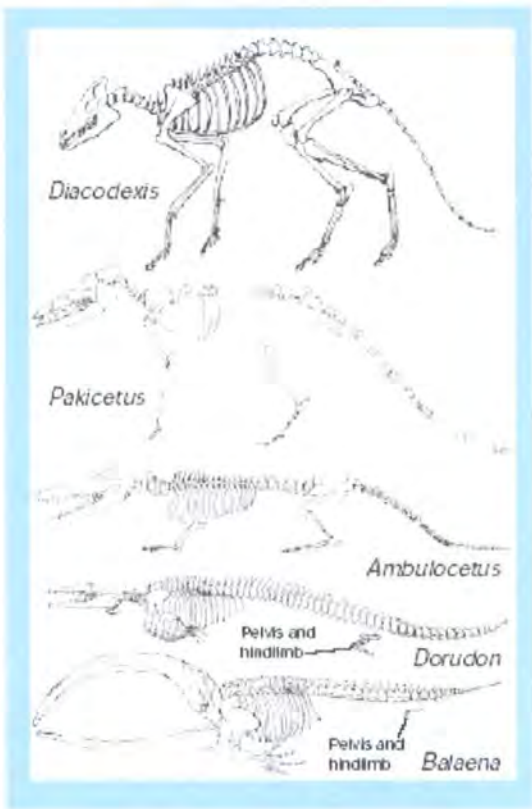
1.1 Adaptations of cetaceans to aquatic life

There are at least one hundred species of marine mammals which depend on the marine environment for most if not all of their life needs (Berta and Sumich 1999). Extant marine mammal species belong to three orders: a) *Carnivora* (the *pinnipeds* such as walruses, sea-lions and seals; polar bears and sea otters), b) *Sirenia* (dugongs and manatees) and c) *Cetacea* (*Mysticeti* such as baleen whales and *Odontoceti* such as dolphins and porpoises). It is commonly accepted that cetacean early evolution started approximately 55 million years ago (Mya), from an artiodactyl land-mammal species (represented by *Elomeryx armatus*) to a semi-aquatic intermediate species (represented by *Rodhocetus balochistanencis*, approximately 45 Mya) and to a fully aquatic species (represented by *Dorudon atrox*, approximately 37 Mya; Thewissen *et al.* 1994; Berta and Sumich 1999; Gingerich 2003; Figure 1.1.1). *Dorodon atrox* belongs to the *Archaeoceti* order, which is comprised by the earliest whales (Thewissen *et al.* 1996). Cetacean adaptation to a fully aquatic existence involved extensive anatomical and physiological modifications.

Early cetaceans had to adapt to four main physical characteristics of the aquatic environment: a) water density, b) pressure, c) thermal conductivity and d) salinity. The main anatomical and physiological features, which were influenced by these parameters, were: a) locomotion, b) diving, c) thermoregulation and d) osmoregulation (see review by Williams and Worthy 2002). However, the physical characteristics of the aquatic environment were not the only challenges that early cetaceans had to face. Studies on several marine pathogenic organisms (including external and internal parasitic organisms) of fish, shellfish and marine mammals have shown that these pathogens

have been associated to the marine environment for long evolutionary periods (see reviews of Howard *et al.* 1983; Lipscomb *et al.* 1994; Kennedy 1999; Higgins 2000).

Figure 1.1.1: Shows the early evolution of cetaceans according to Thewissen *et al.* 1996 (adapted from De Muisson 2001).



For example, the divergence between different cetacean morbillivirus species and a hypothetical common terrestrial ancestor occurred millions of years ago (Barrett *et al.* 1993, 1995; Osterhaus *et al.* 1995). Cetaceans are warm-blooded and breathe air like all mammals; however, they rely on the aquatic environment for their life needs. This may result to an interaction with both the terrestrial and the marine environment pathogen risk.

1.2 MHC Evolution

The immune system in mammals in classified into innate and adaptive (or memory) immunity (see reviews by Roitt *et al.* 2001; Lydyard *et al.* 2004). The innate immune response refers to non-specific immunity such as the phagocytic cells (macrophages, monocytes and neutrophils). The adaptive immune response refers to specific immunity exhibited by lymphocytes (T cells and B cells) which are responsible for the specific recognition of antigens. In addition, the adaptive immune response has also the advantage of 'memory' upon subsequent exposure to a particular antigen. For example, an individual infected by the measles virus will exhibit life-long immunity. The adaptive immune response relies heavily on antigen processing and presentation which is accomplished by MHC molecules.

The Major Histocompatibility Complex (MHC) is a multigene family which codes for cell surface glycoproteins that bind peptides of processed foreign antigens and present them to T-lymphocytes (Klein 1986; Lawlor *et al.* 1990). The MHC plays an important role in the vertebrate immune response, as antigen presentation will subsequently trigger the appropriate immune responses (Klein 1986). MHC genes are classified into two groups, class I and class II. Class I molecules are expressed in most nucleated cells whilst class II molecules are expressed primarily in B cells and macrophages (Klein 1986; Roitt *et al.* 2001). Class I and class II components of MHC were genetically linked when amphibians and mammals diverged from an ancestor 300-350 Mya (Nonaka *et al.* 1997). The gene organization in MHC has been maintained for at least 300 million years in vertebrates (Hughes and Nei 1989, 1990).

1.2.1 MHC function and structure

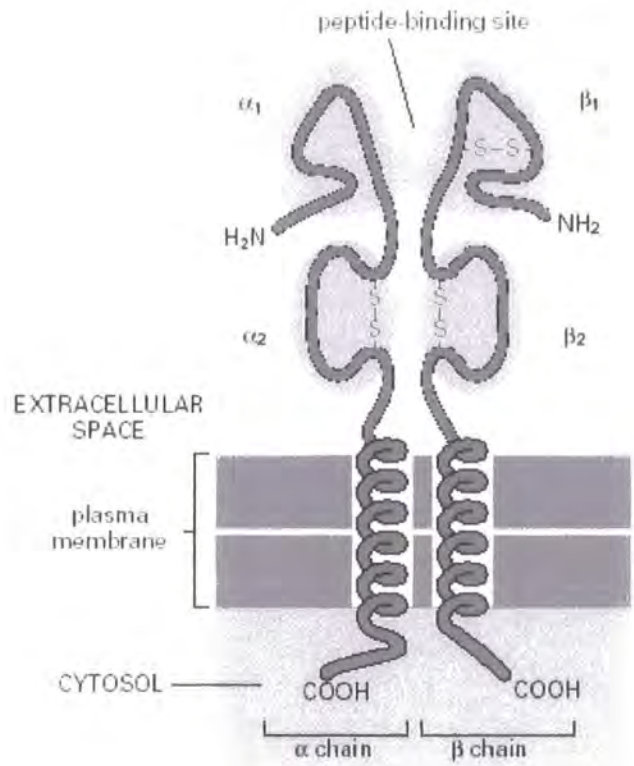
Zinkernagel and Doherty (1974) showed that T-cells will only recognize short fragments of viral proteins when they are bound to specific MHC molecules. In order for the immune response to be initiated, T-cells need to be activated by a specific signal from the MHC molecule and the pathogen peptide. This phenomenon was called MHC restriction, as the immune response is restricted to MHC antigen presentation (Zinkernagel and Doherty 1974). The process of the adaptive immune response begins with the binding of the peptide to the MHC molecule and the presentation of this

complex on the surface of the cell membrane (Klein 1986; Roitt *et al.* 2001; Lydyard *et al.* 2003). Foreign peptides will initiate the cell mediated immune response by triggering the T cells (Roit *et al.* 2001; Lydyard *et al.* 2004). Peptides which are between 13 to 25 amino-acids long and of an extracellular origin will bind to class II molecules and subsequently, they will be presented to helper T-cells (CD4+). Peptides which are between 8 to 11 amino-acids long and of an intracellular origin (cytosolic) will bind to cytotoxic T-cells (CD8+). Helper T-cells induce antibody production and macrophage cells activation whilst cytotoxic cells cause the rupture of the infected cell (Roit *et al.* 2001; Lydyard *et al.* 2003). Therefore, the main function of class I and class II MHC molecules is to present sampled peptides from both the intracellular and the extracellular environment to T cells. It should be mentioned that class II molecules are not restricted to exogenous peptides, for example in a study by Sant (1994) it was shown that endogenous membrane peptides were also presented by class II molecules. There is evidence to suggest that the function of class I and class II molecules overlap. Townsend *et al.* (1989) showed that when cells were bathed in a solution saturated with an extrinsic peptide, both class I and class II molecules bind to the peptide and present it on the cell surface.

Class I and Class II molecules have three different domains: the cytoplasmic, the transmembrane and the extracellurar domain (Figure 1.3.1; Klein 1986). The cytoplasmic domain of the molecule is the part of the alpha chain (class I) and alpha and beta chains (class II), which comes into contact with the cytoplasm, by interacting with the phospholipids of the membrane (Klein 1986). The transmembrane domain of the alpha (class I) and alpha and beta chains (class II) of the molecule contains mainly hydrophobic amino acid residues with a total length of 24 residues (Klein 1986).

The extracellular domain of the MHC molecules has been conserved through approximately 500 Mya of evolution (Klein *et al.* 1997). The extracelular domain of known class I and class II molecules falls into two different groups: the immunoglobin-like domains (immunoglobin fold, *i.e* two sheets of antiparallel β-strands) and the peptide binding domains (MHC fold, *i.e.* α-helix supported by β-strands; Klein and O'hUigin 1993).

Figure 1.2.1: Molecular structure of MHC class II molecule (adapted from Alberts *et al.* 1999).

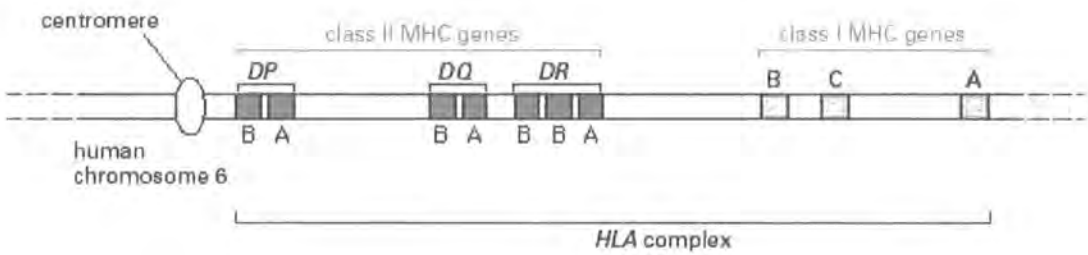


The class I molecule consists a heavy chain (encoded by MHC genes) and an invariant chain (encoded by β2 microglobin gene). The heavy chain is comprised by a pair of α-helixes forming a large groove, which is supported by the eight β-strands. This groove is the binding region of the peptide (PBR) to be presented to T-cells (Figure 1.2.1; Roitt *et al.* 2001). The extracellular domain of the class II molecules comprises of an α-helix and β-strand and together they form the PBR (Figure 1.2.1; Roitt *et al.* 2001).

Genetic organization of Class II molecules (see reviews by Klein 1986; Roitt *et al.* 2001; Lydyard *et al.* 2004)

There are three loci, *DR*, *DQ* and *DP*, which encode the class II molecules and have been identified in several mammalian species (Figure 1.2.2). *DR* is comprised of one α gene and at least nine different β genes, *DQ* is comprised by one functional α gene (*DQA1*) and one functional β gene (*DQB1*) and DP is also comprised by a single functional α gene (*DPA1*) and one functional β gene (*DPB1*). However, *DQ* and *DP* loci also present an additional pair of α and β genes, the functions of which are not well determined (Roitt *et al.* 2001). Class II molecules are heterodimers and are comprised by the products of the α and β genes. For example *DQA1* and *DQB1* encode the HLA-*DQ* antigens. Both α and β genes are comprised of five exons and they have been shown to exhibit high levels of genetic variation in humans. Polymorphism is centered in exon-2 of the α and β gene which codes for the peptide binding region (PBR) of the *DQ* molecule.

Figure 1.2.2: Genetic map of Class I and class II classical loci of the Human Leukocyte Antigens (HLA; the human MHC; adapted from Alberts *et al.* 2002).



Dohorty and Zinkernagel (1975) were the first to show that the product of different MHC alleles binds to a different array of proteins and thus correlating MHC polymorphism to biological function. Furthermore they suggested that, since MHC molecules are co-dominantly expressed, a heterozygote individual will be able to present a wider array of peptides and therefore resist a broader array of pathogens (Dohorty and Zinkernagel 1975). Sequence studies on MHC have shown that certain

class I and class II genes have extremely high levels of polymorphism, among the highest observed in all studied organisms (Klein 1986).

1.2.2 MHC polymorphism

The MHC class I and class II loci have been shown to be highly polymorphic in human, mouse, rat, chicken, bovines, pig, horse, dog and primates (Klein 1986; Udina *et al.* 1994; Trowsdale *et al.* 1995; Ellegren *et al.* 1996; Nasir *et al.* 1997; Wagner *et al.* 1998; Chardon *et al.* 1999; Horin and Matiasovic 2002; Otting *et al.* 2002; Villegas-Castagnasso *et al.* 2003). In humans heterozygosity is as high as 80-90% (Klein 1986). However, polymorphism in MHC genes is not random, as it is mainly restricted to the sites that specify the amino acids of the PBR (Klein and Figueroa 1986). Two of the main reasons that MHC polymorphism has been attributed to positive Darwinian selection are: a) the high number of non-synonymous (*dn*) relative to synonymous (*ds*) rates in the PBR and b) trans-species polymorphism.

Nucleotide substitution pattern

Doherty and Zinkernagel (1975) formed the hypothesis that polymorphism in MHC is maintained by overdominant selection (heterozygote advantage), a suggestion which was further supported by Muruyama and Nei (1981), who showed that mutation and overdominant selection in finite populations are able to maintain the observed high levels of polymorphism observed in MHC. Hughes and Nei (1988, 1989) showed that PBR codons of class I and class II molecules are under overdominant or frequency dependent selection, as the rate of non-synonymous substitution per site (*dn*) was higher than the rate of synonymous substitutions per site (*ds*) in both humans and mice. According to overdominant selection, heterozygotes carrying new mutants have a selective advantage over average individuals (see Chapter 6 for other selection hypotheses; Hughes and Nei, 1988, 1989). Furthermore, MHC variation has been attributed to neutral evolution mechanisms such as the high mutation rate hypothesis (Klein 1978) and the gene conversion or interlocus genetic polymoprhism hypothesis (Lopez de Castro *et al.* 1982). However, according to the neutral hypotheses, the *dn* and *ds* rates are expected to be the same as no selection is involved (Hughes and Nei

1988). Hughes and Nei (1988, 1989) suggested that the higher *dn* rates observed in PBR in MHC do not support neutrality.

Transpecies evolution pattern

Further support for selection was provided from another characteristic of MHC polymorphism, the trans-species pattern of polymorphism (Klein 1980; Klein *et al.* 1998). Alleles in class I and class II have been shown to pre-date speciation events. For example, in humans and chimpanzees certain *DQB1* alleles have persisted since before these two species diverged, 5 to 7 Mya (Lawlor *et al.* 1988; Mayer *et al.* 1988; Gyllensten Erlich 1989). Cluster of trans-species alleles are suggested to represent allelic lineages which were present in the common ancestor of the descendant species (Klein *et al.* 1998). Trans-species polymorphism has been attributed to selection, as neutral polymorphisms are not expected to persist in populations (Klein 1980; Takahata and Nei 1990; Klein *et al.* 1998; De Boer *et al.* 2003).

1.2.3 MHC evolution and pathogens

There are three major advantages pathogens have over their mammalian hosts and these are: a) high replication rate, b) high mutation rate and c) high population numbers (McDade and Worthman 1999). On a population level, a mammal cannot respond to pathogenic pressure as it is relatively long-lived and exhibits relatively low rates of mutation (Du Pasquier 1992). Adaptive immunity through T and B cells compensates for the evolutionary potential difference between pathogens and mammals as their population numbers are high and they also have a high replication rate and a high rate of polymorphism through MHC and cell receptors (Paul 1993; Kuby 1994; Roitt *et al.* 2001; Trowsdale and Perham 2004).

A pathogen plays a very important role in the evolution of the host species. Natural selection has to act on any deleterious effects of those pathogens and the pathogens are evolving to overcome any evolved resistance of the host species. This is commonly referred to as the *arm race* between the immune system of the host and the pathogen. Due to this interaction genetic polymorphism is promoted in host populations for genes involved in the immune response.

1.2.3.1 MHC and infectious diseases

Infectious disease can be caused by a virus, a bacterium, a fungus and a protozoal or a metazoal parasite (Klein and O'Huigin 1994). Infection may be acute or chronic. An acute infection has a short duration and it is strongly pronounced while a chronic infection has a long duration and it may or may not be pronounced (Klein and O'Huigin 1994; Roitt *et al.* 2001; Lydyard *et al.* 2004). Klein and O'Huigin (1994) have divided pathogenic organisms into three main categories in the context of MHC evolution: a) pathogens causing acute infection, exhibiting high virulence, b) pathogens causing chronic infection and disease and c) pathogens causing chronic infection and causing disease only under certain conditions.

Highly virulent pathogens such as influenza virus, smallpox virus, rubella virus and parasites such as *Trypanosoma gambiense* and *Plasmodium falciparum* fall into the first category (Klein and O'Huigin 1994). However, the high virulence and the fact that these pathogens have been associated with human hosts for the last 10,000 years suggest that pathogens of the first category exert a relatively small amount of selection on MHC (Klein and O'Huigin 1994). The high virulence immobilizes the immune system and the host is killed quickly thus not allowing selection to act on the immune response (Klein and O'Huigin 1994). Furthermore, the HLA polymorphism now present in the global population is more than 10000 years old, as the time needed for a novel variant to incorporate as a polymorphism into the global population is longer than 0.5 Mya (Klein 1991).

The second category of parasites follow the Fahrenholz's rule, *i.e* pathogens of this category have speciated in parallel with the hosts and exhibit similar phylogenetic histories (Soeda *et al.* 1980; Shaden and Villareal 1993). It is suggested that the immune system of the host, including MHC molecules, has co-evolved with the pathogen (Klein and O'Huigin 1994). Pathogens such as papova virus, polyoma virus, schistosomata worms and non-falsiparan plasmodium species have been suggested to fall into the second category (Klein and O'Huigin 1994). These pathogens may be a source of selection pressure (Klein and O'Huigin 1994).

The third category of pathogens is that of organisms often referred to as opportunistic pathogens, as they are well adapted to their hosts and normally they are not pronounced (Klein and O'Huigin 1994). However, disease is caused under

favorable conditions, such as when the host is immunogically compromised (Klein and O'Huigin 1994). Organisms which fall into this category are diverse, including *cytomegalovirus*, herpes simplex virus, *Escherichia coli*, *Pseudomonas sp.*, *Streptococci sp.*, Candida *sp.* and *Pneumocystis carini* (protozoan; Klein and O'Huigin 1994). These pathogens are ubiquitous in nature and have been associated with their host for long evolutionary periods, and they have been suggested to be the main selection forces in the evolution of the vertebrate immune response (Klein and O'Huigin 1994). Trans-species evolution and substitution patterns exhibited by MHC polymorphism are suggested to be influenced and maintained by the pathogens of the second and third category, as these pathogens have provided a continuous selection pressure on MHC genes to diversify and for variants to endure (Klein and O'Huigin 1994).

1.3 Marine pathogen environment of Cetaceans

The marine pathogen environment is poorly known and understood; however, in the last decade herpesvirus, poxvirus, adenovirus, calicivirus, papilomavirus and morbillivirus were all identified in marine mammals around the world (vanBressem *et al.* 1999). In the last 20 years thousands of marine mammals have died due to epizootics and viral infections (vanBressem *et al.* 1999).

Furthermore, Goyal *et al.* (1983) and Griffin *et al.* (2003) demonstrated that human enteric viruses from fecal matter remain virulent in the marine environment for up to 17 months, thus illustrating the occurrence of human pathogen spillover from the terrestrial environment to the marine (Bronwell *et al.* 2000). The Canine Distemper Virus spillover from dogs, which took place in 1986 in lake Baikal in Russia, had devastating effects on the fresh water Baikal seal population (Mamaev *et al.* 1996; Forsyth *et al.* 1998). Morbilliviruses infections are well documented, as they cause serious disease with a high mortality rate and they persist in the marine environment. It has been indicated that morbilliviruses species, such as Phocine Distemper Virus (PDV), Porpoise Morbillivirus (PMV), Dolphin Morbillivirus (DVM) and Cetacean Morbillivirus (CMV) have been present in the marine environment for an extensive period of time, long enough to determine through molecular data that marine

morbilliviruses have diverged from the terrestrial species long ago (Domingo *et al.* 1995; Duignan *et al.* 1995; Rima *et al.* 1995).

1.3.1 Marine pathogenic organisms

1.3.1.1 Viral diseases

There have been five main epizootic incidents in cetaceans involving morbillivirus infections of five different species that have been documented in the literature. The first was in 1987-88 on the east coast of the United States, where 50% of the in-shore populations of bottlenose dolphins (*Tursiops truncatus*) died (Limpscomb *et al.* 1994). The second was in 1988-90 on the North-Eastern Atlantic affecting the harbor porpoise (*Phocoena phocoena*; Visser *et al.* 1993). The third was in 1990–92 on the Mediterranean affecting the striped dolphin (*Stenella coeruleoalba*; Di Guardo *et al.* 1995). The forth was in 1993-94 affecting the bottlenose dolphin in the Gulf of Mexico (Duignan *et al.* 1995). Recently there have been reports of morbilliviral infection in the Pacific in the common dolphin (*Delphinus delphis*; Van Bressem *et al.* 2001).

The pathogenic agents were identified using monoclonal antibodies and sequence analysis to be Dolphin Morbillivirus (DMV) and Porpoise Morbillivirus (PMV), two distinct species but closely related (Rima *et al.* 1995). In 2000 a novel morbillivirus species, Pilot Whale Morbillivirus (PWMV), was identified in a long-finned pilot whale (*Globicephalus melas*) distinct from, but closely related to DMV and PMV (Taubenberger *et al.* 2000). However, the origin of these cetacean morbilliviruses and the mechanism of dispersal still remain unknown (Wild *et al.* 1995; Van Bressem *et al.* 2001). There have been nine families of viruses identified in cetacean around the world (Van Bressem *et al.* 2001; Table 1.3.1). Three distinct families, *Paramyxoviridae*, *Poxviridae* and *Papovaviridae* pose a serious risk and they have had an affect on increasing mortality rates (*Paramyxoviridae* and *Poxviridae*; see Table 1.2.1) as well as inhibiting copulation and reproduction in several cetacean species (*Papovaviridae*; see Table 1.3.1; Van Bressem *et al.* 1999; Van Bressem *et al.* 2001).

1.3.1.2 Bacterial diseases

The organisms responsible for infections in cetaceans are the same agents as in terrestrial infections. Species such as *Erysipelothrix rhusiopathiae*, *Staphylococcus*

aureus, *Streptococcus*, *Pseudomonas aeruginosa*, *Escherichia coli* and *Vibrio alginolyticus* have been found to cause disease in cetaceans from around the word (Higgins 2000; Lipscomb 2000; Van Bressem *et al.* 2001; Table 1.3.2).

Table 1.3.1: Viral agents in Cetacean species.

FAMILY	NAME	SPECIES
Adenoviridae		*B. borealis,. B. mysticetus, D. leucas*
Caliciviridae	San Miguel sea lion virus, Swine virus, Cetacean calicivirus	*T. truncatus, E. robustus, B. borealis, B. physalus, B. mysticetus, P. macrocephalus*
Hepadnaviridae		*L. obliquidens*
Herpesviridae	Herpes-like viruses, Porpoise herspervirus	*D. leucas, L. obscurus* *P. phocoena*
Orthomyxoviridae	A/Whale/Maine/1/84, A/Whale/Maine/2/84, A/Whale/PO/19/76	*G. melas, Balaenopteridae*
Rhabdoviridae	Dolphin rhabdovirus like virus	*L. albirostris, D. delphis, G. melas, S. coerulealba, P. phocoena, T. truncatus,P. crassidens*
Papovaviridae	Phocoena spinipinnis papil-lomaviruses, Papillomaviruses-like viruses	*P. spinipinnis, L. obscurus, P. phocoena*
Paramyxovirida	Cetacean morbilliviruses	*P. phocoena, K. breviceps, S. coeruleoabla, S.frontalis, L. obscurus, L. acutus, L. albirostris, G. griseus, D. capensis, D. delphis, T. truncates, D. del. Ponticus, G. melas, G. macrorhynchus, P. crassidens,F. attenuata, B. physalus, P. acutorostrata*
Poxviridae	Cetacean poxviruses	*T. truncatus,L. acutus, L. obscurus, D. capensis, O. orca, P. spinipinnis*

Furthermore, a novel *Brucella* species has been identified to be common in a number of cetacean species although abortion caused by *Brucella* infection has been documented only in bottlenose dolphins (vanBressem *et al.* 2001). A novel species of *Staphylococcus* (*Staphylococcus delphini*) isolated by suppurative skin lesions (Lipscomb 2000) has also been reported. The zoonotic potential of bacterial diseases

has been demonstrated to occur in an amphidromus direction between the marine and the terrestrial environment; there are documented incidents of humans who work at close proximity to captive dolphins that exhibit bacterial infections developing localized infections of the same type (Palmer *et al.* 1991; Lipscomb 2000).

Table 1.3.2: Bacterial agents in cetacean species.

BACTERIAL SPECIES	CETACEAN SPECIES
Erysipelothrix rhusiopathiae	*T. truncatus, S. plagiodon, G. grampus, L. obliquidens, T. aduncus, L. albirostris*
Staphylococcus aureus	*T. truncatus, T. gilli,*
Staphylococcus delphini	*T. truncatus*
Streptococcus spp.	*T. truncatus, Globicephala melas*
Pseudomonas spp.	*T. truncatus, O. orca, D. leucas, D. delphis, L. obliquidens*
Clostridium spp.	*T. truncatus, D. leucas*
Nocardia spp.	*T. truncatus, T. aduncus, O. orca, S. longirostris, P. phocoena, K. breviceps, G. griseus*
Vibrio alginolyticus	*L. acutus, T. truncatus*
Escherichia coli	*T. truncatus, D. leucas,*
Brucella spp.	*T. truncatus, B. acutorostrata, Delphinus delphis, L. acutus, B. acutorostrata, B. bonaerensis*
Salmonella spp.	*T. truncatus, Orcunus orca, Globicephala spp.*
Actinobacillus delphinicola	*Phocoena phocoena, S. coeruleoalba, Mesoplodon bidens*
Vibrio spp.	*D. leucas, T. truncatus, T. aduncus,*

1.3.1.3 Protozoal diseases

Toxoplasmosis, which is caused by *Toxoplasma gondii,* is the main protozoal infection found in cetaceans and it has also been observed in the majority (86%) of individuals for which the cause of death was a morbillivus infection (Lipscomb 2000; Mikaelian *et al.* 2000). It was first observed in a mother and a calf bottlenose dolphin, and it was initially believed to be a unique and isolated incidence, as Toxoplasma does

not survive in marine fish and invertebrates (Higgins 2000; Lipscomb 2000). Further evidence came from a stranded spinner dolphin (*Stenella longirostris*) and from four striped dolphins that died during the morbillivirus epizootic (Lipscomb 2000). This association is also observed in dogs infected with Canine Distemper Virus (CDV), possibly caused by morbillivirus-induced immunosuppresion (Barret 1999; Lipscomb 2000).

Ciliated protozoa have also been identified within areas of dermatitis in Atlantic bottlenose dolphins (Higgins 2000; Lipscomb 2000). It is suggested that these organisms are secondary invaders of wounds inflicted by other causes (Lipscomb 2000).

1.3.1.4 Fungal diseases

Fungal diseases have been identified mainly in captive cetaceans (Higgins 2000; Lipscomb 2000). Species such as *Laboa laboi*, *Candida spp.*, *Aspergilus spp.*, *Sporothrix schenckii*, *Blastomyces dermatitidis* and *Cryptococcus neoformans* (in a free-living striped dolphin) were identified as disease causing agents (Higgins 2000; Lipscomb 2000).

1.3.2 Anthropogenic factors and infections

Pollution, fisheries interaction, habitat loss and habitat degradation are the main anthropogenic factors which have a direct affect on populations of cetacean species. These factors have the potential to aggravate the infection process and the clinical pathology of a pathogenic organism to a given individual, and they may prevent the establishment of an enzootic infection, thus favoring recurrent epizootics (Van Bressem *et al.* 2001). This is accomplished through the reduction of the number of animals in a population and therefore reducing the number of susceptible animals (Van Bressem *et al.* 2001).

1.3.2.1 Pollution

Environmental contaminants present a high risk in populations of marine mammals as they are long lived, predators at the top of the food chain and have large energy reserves of subcutaneous fat – blubber (Binh Minh *et al.* 1999; Becker 2000; Taddel *et al.* 2001). The majority of contaminants are polychlorins, which are

lipophylic (fat-soluble) chemicals, and this is the main reason why they become increasingly accumulated at each level in the food chain as fat is metabolized but not the contaminants (Ross *et al.* 1996; Becker *et al.* 1997). It has been demonstrated that marine mammals accumulate high concentrations of environmental contaminants such as Polychlorinated biphenyls (PCBs), DDT and its metabolites, Polychlorinated Dibenzo-*p*-Dioxins (PCDDs) and Polychlorinated Dibenzofurans (PCDFs). This is achieved through inhabitation of polluted areas and/or through accumulation in the food chain (Ross *et al.* 1996; Becker *et al.* 1997; Binh Minh *et al.* 1999; Becker 2000; Taddel *et al.* 2001). These chemicals interfere with the homeostasis of marine mammals, as they mimic hormones (Ross *et al.* 1996).

The immune system has been shown to be particularly sensitive to the toxic actions of the aforementioned contaminants. Osterhaus *et al.* (1995) demonstrated that seals which fed on contaminated fish from the Baltic Sea showed changes in several immune functions including impairment of *in vitro* killer and T cell responses and *in vivo* delayed type hypersensitivity and antigen-specific B cell responses. This was the first time that it was shown experimentally that pollution can cause immune suppression through impairment of immune functions in marine mammals. Further observations by Hall *et al.* (1992) and Aguilar *et al.* (1994) indicated that animals which succumbed to morbillivirus epizootics had higher concentrations of organochlorides than the live individuals at the given time period.

These observations do not provide solid evidence for contaminant initiated immunotoxicity and thus the complete implication of pollution and reoccurring epizootics and viral infections has not been determined. However, it is suggested that pollution creates conditions that favor the emergence of infections in cetaceans (Van Bressem *et al.* 1999; Van Bressem *et al.* 2001).

1.3.2.2 Fisheries interaction

It is estimated that about 800 cetaceans die every day by fisheries interactions (Fertl 2002). Fishing gear such as gillnets, set nets, trammel nets, seines, trawling nets and longlines pose the greatest danger (Fertl 2002; Read *et al.* 2004). The total number is 308,000 cetaceans annually and this has an effect on the population dynamics of cetaceans and infectious disease agents (Read *et al.* 2004).

A long period of exposure to a given pathogen is required, in order for enzootic infections to occur and for the hosts to develop immunity (vanBressem *et al.* 1999). Fisheries remove a large number of individuals from cetacean population and this constantly disrupts the balance between the host immune response and pathogen population structure; by constantly removing individuals from a population epizootic outbreaks are favored (Van Bressem *et al.* 1999; Vossen *et al.* 2002).

1.4 Objectives

Several studies on MHC polymorphism have indicated that the evolution of diversity in immune system genes responsible for antigen presentation in the terrestrial mammals studied is driven by selection (Klein 1986). The terrestrial pathogen environment has been suggested as the primary selection pressure on MHC genes to diversify and for polymorphism to be maintained (Klein and O'Huigin 1994).

The marine pathogen environment exhibits a degree of similarity with the terrestrial environment. Marine pathogenic organisms, which fall into the second and third category of Klein and O'huigin (1994) classification, have been identified in cetaceans (see table 1.3.1). For example, morbillivirus and influenza virus infectious are highly virulent and may be classifies in the first category of pathogens, whilst, viruses of the adenoviridae, caliciviridae, herpesviridae, papoviridae families may be classified in the second category. In addition, organisms such as *streptococci sp.*, *pseudomonas sp.* and *escherichia coli*, which have been classified in the third category, are also found in cetaceans (see Table 1.3.2). Furthermore, studies on marine mammals have shown that immunosuppresion due to pollution is observed in pinnipeds and cetaceans and this has also been observed in terrestrial mammals, confirming that there are aspects of the immune response in cetaceans similar to terrestrial mammals (De Guise *et al.* 1997). However, it is not clear whether cetaceans will experience the same selective pressures as terrestrial mammals.

The main objective of this study is to evaluate the role of the pathogen environment on the evolution of the exon-2 of the class II *DQB1* MHC locus in order to test the hypothesis that selection generates and maintains variation at this locus. The evolution pattern and the role of natural selection on exon-2 *DQB1* locus in cetaceans is

assessed by phylogenetic and distance analyses at three different levels: I) order level (Chapter 3), II) suborder and species level (Chapter 4) and III) population level (Chapter 5). The hypotheses tested in this study are:

I) The marine pathogen environment presents comparable selection pressures to those of the terrestrial environment. In Chapter 3, the level and pattern of *DQB1* locus variation in cetaceans is assessed and compared to that of terrestrial mammals. This analysis will evaluate the role and magnitude of selection on *DQB1* evolution on an extensive sample of cetacean species.

II) Social behaviour and structure will affect pathogenic pressure. In Chapter 4, the effects of social behaviour and social structure on *DQB1* evolution are assessed through the comparison between *Odontoceti* and *Mysticeti* suborder and species. This analysis will investigate whether environmental differences in cetaceans may also affect the strength of selection pressure on *DQB1* locus.

III) Populations of cetacean species with a world-wide distribution across different habitats and geographic regions are under differential selective pressure. In Chapter 5, the population structure according to *DQB1* locus is assessed, in order to evaluate the role of selection across different environments and geographical scales in cetacean populations. This analysis will investigate the role of selection versus other evolutionary forces (such as genetic drift) in shaping *DQB1* variation in cetaceans.

CHAPTER 2:

MATERIALS AND METHODS

2.1 CHEMICALS

All chemicals were either Analar or "molecular biology" grade and were purchased from Sigma-Aldrich (Poole, Dorset, UK) or Merck Chemicals Ltd. (Poole, Dorset, UK) unless otherwise stated. Molecular biology kits were purchased from Sigma-Aldrich or Promega Ltd. (Chailworth science park, Southampton, UK) or Qiagen (Crawley, West Sussex, UK).

2.2 SAMPLES

Tissue samples were collected over a number of years from live or stranded dead individuals from 18 different cetacean species (Chapter 3) and provided by Dr A. R Hoelzel (Molecular ecology and evolution group, University of Durham, UK).

2.3 EXPERIMENTAL PRECAUTIONS

All glassware, buffers (where possible) and solutions were autoclaved at 121°C for 20 minutes to effect sterilization. If autoclaving was unsuitable, solutions were sterilized by passage through a 0.2μm syringe filter. Solutions used in PCR (see 2.5.1) method were made to a final volume, sterilized by passage through the 0.2μm filter, 30 minutes autoclaving and exposed to short-wave UV light prior to usage.

All work involving PCR and bacteria laboratory techniques was performed in a sterile environment under conditions of good microbial/ sterile culture practice. DNA-AWAY solution (cat. no: 7010; Merck Chemicals) was used to treat working area and equipment prior to PCR work. In addition, all Gilson Pipetman® pipettes were routinely treated with DNA-AWAY solution and exposed to UV light. Glassware equipment used in SSCP method were soaked for a minimum of 30 minutes in sterile water then washed and treated with dimethyl-dichlorosilane (D-3879; Sigma-Aldrich)

2.4 ROUTINE TECHNIQUES

All the techniques described in this section are routinely used in the laboratory and are performed in conjunction with many of the more specific protocols described later in this chapter.

2.4.1 DNA routine techniques

2.4.1.1 Spectrophotometric quantification of DNA

The concentration and quality of the DNA was determined by measuring the optical density at 260 and 280nm. We considered that at 260nm the OD 1 corresponds to 50μg of double stranded DNA. The reading at 280nm accounts for protein contamination of the sample. The ratio of absorbance at 260 and 280nm gives a measure of the purity of the DNA. A pure DNA solution should have a ratio of 1.8 (Sambrook *et al.* 1989).

2.4.1.2 Restriction endonuclease digestion

This technique was routinely used in order to isolate PCR insert from plasmid DNA. In a microfuge tube, 0.2-5 μg of plasmid DNA, 0.5 μl of each the appropriate restriction enzymes (Promega), 2 μl of multicore restriction buffer (Promega) and distilled water to final volume (20 μl) were mixed and incubated for 1 hour at 37°C. The results of enzymatic restriction were subsequently analysed by agarose gel electrophoresis.

2.4.1.3 DNA gel electrophoresis

Agarose gel electrophoresis was used to isolate DNA fragments produced by DNA isolation or PCR method or linearised plasmids produced by enzymatic restriction. The DNA fragments were separated on the basis of their size and charge. The fragments were then assessed by comparison with standard DNA ladders (L-52721; Sigma-Aldrich) run in parallel. DNA samples were routinely electrophoresed in 1.5% agarose gel. The electrophoresis was carried out using 0.5x Tris Borate EDTA Buffer (TBE; 10X TBE: 50mM Na4EDTA, 10 mM EDTA, 0.4 M Tris boric acid pH 7.5).

The appropriate weight of agarose was melted in 1X buffer, left to cool for a few minutes, 0.68 µg/ml ethidium bromide added and the mixture poured in a Bio Rad-Wide Mini-Sub™ horizontal gel tray with a comb to provide wells into which the DNA samples could be subsequently loaded. The gel was then allowed to set at room temperature before addition of the running buffer in the gel tank, removal of the comb and loading of the samples. Two or five (PCR product or plasmid respectively) µl of DNA were combined to 2 µl of DNA loading dye (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, 30% (v/v) glycerol, 6mM EDTA) in each sample before being loaded into separate well of the agarose gel.

Unless otherwise stated, electrophoresis was carried out for 30 minutes at 100V. The gel was then viewed under short-wave UV light. The ethidium bromide intercalates into the DNA which fluoresces under UV light thus making the DNA bands visible. Gels were photographed using a Bio Rad Gel Doc 2000 gel documentation system with Quantity-One™ 4.0.3 software.

2.4.1.4 DNA sequencing reactions

DNA, prior to sequencing, was purified using a Qiagen QIAquick PCR Purification Kit™ to remove primer dimers, unincorporated dNTPs, and unwanted chemicals. Occassionally, non-specific amplification products were obtained in addition to the expected product. In these cases, the Qiagen Gel Extraction Kit™ was used to purify only the desired PCR product. The instructions provided with the kits were followed. The purified DNA was then sequenced by DBS Genomics (University of Durham) using the ABI Big-Dye Terminator sequencing kit™ and running of the products on an ABI Prism 373 or 377 polyacrylamide slab gel automated sequencer. Each reaction was carried out using either universal primers (M13) for plasmid DNA or PCR specific primers for PCR fragments, as appropriate.

2.4.2 Bacterial routine techniques

2.4.2.1 Bacterial strains

Escherichia Coli (*E. Coli*) XL1 blue [*rec*A1, *end*A1, *gyr*A96, *thi-1*, *hsd*R17 (r_K^-, m_K^+), *sup*E44, *rel*A1, λ^-, *lac*$^-$, [F' *pro*AB, *lac*I^qZΔ *M15*, Tn10(Tet^r)]] strain was used for transformation of the plasmids containing the PCR product, generated during this work and were freshly made competent prior to usage.

2.4.2.2 Preparation of competent *E. Coli*

Fresh competent *E. Coli* were prepared by heat shock using a modified version of Chung and Miller (1988) method. Host strain *E. Coli* were grown overnight on Luria-Bertani (LB) agar (1% w/v Bacto-tryptone, 0.5% w/v Bacto-yeast extract, 1% w/v NaCl and 1.5% w/v Bacto-agar) at 37°C then a single colony picked and grown in 20 ml LB broth (1% Bacto-tryptone, 0.5% Bacto-yeast and 1% NaCl) in a shaking incubator at 37°C to a cell density of approximately $5x10^7$ cells/ml (corresponding to a final O.D. of 0.5 at 600 nm). Cultures were then cooled on ice before centrifugation in two aliquots of 10 ml for 10 minutes at 1000 g at 4°C. Each cell pellet was resuspended in 1 ml of LB containing 10% PEG (v/v; MW=3.350), 5% (v/v) DMSO, 10 mM $MgCl_2$ and 10 mM $MgSO_4$. After a further 10 minutes on ice, cells were divided into 100 µl aliquots. Cells were then ready for transformation.

2.4.2.3 Transformation of competent cells

Five microlitres of the ligation reaction were added to an aliquot of the competent cells and incubated on ice for 30 minutes. Three controls were used: i) an aliquot which contained 5 µl of sterile water, ii) an aliquot which contained 0.5 µl of T-Vector and iii) an aliquot which contained 10 pg of blue-script plasmid (Promega). The samples were then heat shocked at 42°C for 90 seconds; thereafter 1 ml of LB broth containing 25mM of glucose was added and the samples incubated for 1 hour at 37°C. Samples were plated on LB agar containing 12.5 µg/ml tetracycline, 100 µg/ml ampicillin, 40 µg/ml X-gal and 0.1mM IPTG (all provided by Sigma-Aldrich). Transformants were recognized by their ability to grow on the agar and disrupting the *lacZ* gene (exhibiting white colonies). A single colony was then picked and grown overnight at 37°C in LB broth containing antibiotic for selection. Bacteria were stored thereafter with 15% glycerol at –80°C.

2.4.2.4 Preparation of plasmid DNA

Small scale preparation of plasmid DNA were performed using the Gene-elute miniprep kit™ procedure (Sigma-Aldrich). This method is based on alkaline lysis of bacterial cells followed by adsorption of DNA onto silica in the presence of high salt (Birnboim and Doly 1979).

A single bacterial colony was used to inoculate 5 ml of LB broth containing the appropriate antibiotic (50 μg/ml ampicillin), then incubated at 37°C overnight with vigorous shaking at 120 rpm. A 1.5 ml aliquot of this overnight culture was pelleted by microcentrifugation at 13000 rpm for 1 minute. The supernatant fluid was discarded and the cell pellet was processed following the kit instruction manual. The Gene-elute miniprep procedure uses the alkaline lysis method of Birboim and Doly (Birnboim and Doly 1979) as modified by Sambrook *et al.* (1989). Bacteria were lysed under alkaline conditions, and the lysate is subsequently neutralised and adjusted to high-salt binding conditions in one step, ready for purification on the Gene-elute silica-gel membrane. The DNA is finally eluted in 50 μl of sterile water and stored at –20°C.

2.5 MOLECULAR GENETIC METHODS

2.5.1 Polymerase Chain Reaction (PCR) method

DNA sequencing of the PBR region of *DQB1* gene was carried out by enzymatic amplification, followed by cloning and sequencing of the final products. The exon-2 PBR region was amplified using the primers (CTGGTAGTTGTGTCTGCACAC and CATGTGCTACTTCACCAACGG) developed by Tsuji *et al.* (1992). The reaction conditions were 10mM Tris HCL (p.H=8.3), 50mM KCL, 2.5mM $MgCL_2$, 0.2mM of each dNTP, 0.25 μM of each primer, 2 units of *Pfu Taq* polymerase (Promega) and 100-150ng of template DNA in a 25 μl final volume. Enzymatic amplifications were performed on either a Hybaid PCR-EXPRESS™ or MJ Research-INC PTC-100™ PCR thermocycler. Thermal cycling involved 35 cycles at 94^0 C for 1 min., 55^0 C for 2 min. and 72^0 C for 2 min. The products of the reaction were cloned to pGEM T-vector system (Promega).

2.5.1.1 Cloning of PCR products

The PCR products generated during this work, when necessary, were cloned using Easy T-Vector Cloning kit (Promega) according to the manufacturer instructions. The PCR fragments were purified using Quigen PCR purification kit™ and inserted into the *Eco*R1 site of the pGEM®-T vector plasmid. Since the PCR fragments were generated using *Pfu* DNA polymerase (Promega) they were blunt ended and therefore an A-Tail reaction was required. Five microlitres of the PCR product, 2 μl of 10x *Taq*

polymerase buffer, 0.6 µl of 2mM $MgCl_2$, 1 µl of 10mM ATP and 0.25 µl of *Taq* polymerase (Promega) were mixed and incubated at 72°C for 30 minutes.

The ligation reaction was carried out using an insert-to-vector ratio ranging from 3:1 to 10:1. In a microfuge tube 0.5 µl of the pGEM T-Vector (10 ng/µl), 5 µl of 2X Rapid Ligation Buffer, the appropriate amount of PCR product, 1 µl of T4 DNA ligase (4 U/µl) and distilled water to a volume of 10 µl were mixed and incubated for 25 minutes at room temperature. The ligation reaction was then incubated at 4°C o/n and subsequently stored at -20°C until use for transformation into the competent cells.

2.5.2 Single Stranded Conformation Polymorphism (SSCP) method

Two microlitres of denaturing loading buffer [95% (v/v) Formamide, 0.1% (w/v) Bromophemol blue, 0.1% (w/v) xylene cyanol and 10mM NaOH; Sigma-Aldrich] were added to 2 µl of PCR product, mixed thoroughly and denatured for 2 min on either a Hybaid PCR-EXPRESS™ or MJ Research-INC PTC-100™ PCR thermocycler. Following the denaturation step, the single stranded samples were immediately placed on ice for a minimum of 2 minutes and were loaded on a non-denaturing acrylamide gel [6% (v/v) 49:1 acrylamide: bis-acrylamide, 10% (v/v) glycerol and 1x TBE] for 6 h and 40 W migrations at 4°C. The gel was incubated for 20 min. with the fluorescent GelStar™ Nucleic Acid Gel Stain (BioWhittaker, Rockland, ME, USA) according to manufacturer instructions. Allelic conformation was visualized by exposure to short-wave UV light and photographed using a Bio Rad Gel Doc 2000 gel documentation system with Quantity One 4.0.3 software. The allelic patterns for each individual where scored and genotypes assigned. The PCR product of individuals exhibiting unique patterns of allelic variation was investigated by cloning (see 2.5.1).

In order to exclude the possibility of *E. coli* recombinant artifacts by mismatch repair of heteroduplexes (Longeri *et al.* 2002) the following procedure was performed. For each of the unique genotypes, twenty-four recombinant clones were purified (see 2.4.2) and insert DNA was isolated by restriction endonuclease digestion (see 2.4.1). Four microlitres of the digestion mixture (containing 50-80ng of insert DNA) were mixed to 2 µl of denaturing loading buffer and loaded on a non-denaturing acrylamide gel according to the procedure mentioned previously. Ten clones from each unique genotype were sequenced in the reverse and forward directions in order to obtain identical clones form each unique PBR sequence, from different amplifications, in order

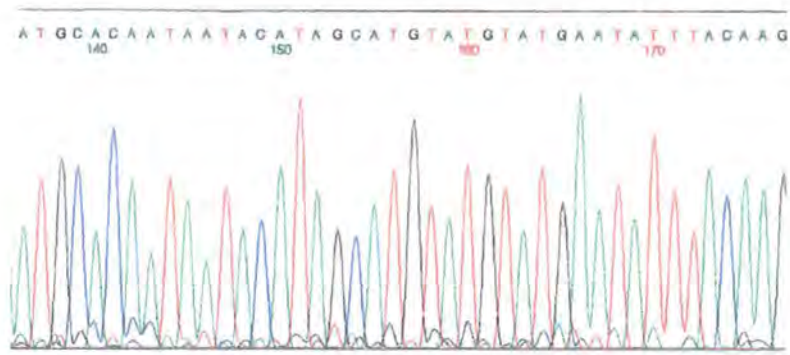
to confirm the nucleotide sequence. The files of the sequencing results were processed by Sequencer™ version 4.1.2 software program (Figure 2.5.2).

Figure 2.5.1: SSCP profile of 24 recombinant clones of a heterozygous individual (*O. orca*). Alleles (*O. orca-1* and *O. orca-2*) exhibit a single point difference (A to G) on site 160 (see Chapter 3, Figure 3.2.1). Double-stranded DNA is indicated by DS.

2.5.3 Sequence alignment analysis

A fragment of 240 bp was successfully isolated through PCR amplification; however, due to difficulties interpreting the beginning and ending flanking regions of the PCR amplicon, 171 bp were used in the analysis. The sequences from different individuals and different amplification reactions were aligned by using Sequence™ software. Multiple alignments and their corresponding chromatograms (Figure 2.5.1) were edited by eye for discrepancies in mistaken nucleotide assignment by automated sequencing method. BLAST version 2.2.9 (http://www.ncbi.nlm.nih.gov/BLAST; Zhang and Modden 1997) was used in order to confirm that DNA sequences were the result of the amplification of exon-2 *DQB1* locus. Both nucleotide and amino acid alignments illustrated the conservation of motifs between cetacean sequences identified in this study and exon-2 *DQB1* sequences available in the BLAST database.

Figure 2.5.2: Portion of a sequence chromatogram resulting from automated sequencing. The sequence is represented in text at the top of the figure (A=Adenine, C=Cytosine, G=Guanine, T=Thymine). The numbers under the text indicate the number of the base sequentially within the chromatogram.



In addition, the resulting sequences of the present study were aligned with 6 unique cetacean sequences (labeled by *–c*) provided by Murray *et al.* (1995) and 27 unique cetacean sequences (labeled by *–j*) provided by Hayashi *et al.* (2003). This combination of sequences included all exon-2 *DQB1* unique alleles identified to date in cetacean species and it was used in the genetic and phylogenetic analyses throughout the present study. The sequence alignment generated by Sequencer™ can be exported as a consensus file and used for further genetic analyses.

2.6 GENETIC ANALYSES

2.6.1 Nucleotide substitution analysis

Nucleotide substitutions in protein coding DNA sequences are divided into two classes: synonymous and non-synonymous (Kumar *et al.* 1993). Synonymous substitutions are nucleotide substitutions which do not result in amino acid change while non-synonymous substitutions are those which result in a different amino acid. Since the majority of non-synonymous substitutions are subject to purifying selection, the rate of synonymous substitutions (*ds*) is higher than the rate of non-synonymous substitutions (*dn*; Miyata *et al.* 1980; Kimura 1983).

However, under certain conditions non-synonymous substitution rates may be accelerated by positive selection (see review by Hughes 1999). The *dn*/*ds* ratio (ω) is used in order to assess the adaptive molecular evolution between DNA sequences of a

given gene, as ω measures the difference between the two rates (Hughes 1999; Yang and Bielawski 2000). For example, a neutral amino acid difference will be fixed at the same rate as a synonymous substitution and thus ω=1; a deleterious amino acid difference, however, will be removed by purifying selection reducing the fixation rate of the non-synonymous substitution, resulting in ω<1. Therefore, only an advantageous non-synonymous substitution will offer a selective advantage and will be fixed in a higher rate than a synonymous substitution resulting in ω>1. There are several methods for estimating the *dn* and *ds* values between protein coding DNA sequences. For the purpose of the present study I will follow the classification by Yang and Bielawski (2000) according to whom, estimating methods are categorized into two classes: a) approximate methods and b) maximum likelihood methods.

2.6.1.1 Approximate methods

In the approximate methods, the number of synonymous (*S*) and non-synonymous (*N*) sites, *i.e.* the sites at which synonymous and non-synonymous substitutions potentially occur, are first computed (Nei and Gojobori 1986; Li *et al.* 1985; Li 1993; Pamilo and Bianchi 1993). The numbers of synonymous (S_d) and non-synonymous (N_d) that have occurred between pairs of sequences are computed, considering all pathways of nucleotide substitution between each pair of codons compared (Nei and Gojobori 1986; Li *et al.* 1985; Li 1993).

Nei-Gojobori method

The Nei-Gojobori method (Nei and Gojobori 1985) uses the quantities of *S*, *N*, S_d and N_d in order to calculate the proportion of synonymous (p_s) and non-synonymous (p_n) nucleotide differences per synonymous and non-synonymous site respectively: $p_s = S_d/S$ and $p_n = N_d/N$ (Nei and Gojobori 1985). The average estimate of the number of synonymous substitutions per synonymous site and non-synonymous substitutions per non-synonymous sites can be obtained by applying the formula of Jukes-Cantor:

$$ds = -\tfrac{3}{4} \ln(1-\tfrac{4}{3} p_s) \text{ with variance } V(ds) = p_s (1-p_s)/[1-(1-\tfrac{4}{3} p_s)^2 S],$$

$$dn = -\tfrac{3}{4} \ln(1-\tfrac{4}{3} p_n) \text{ with variance } V(dn) = p_n (1-p_n)/[1-(1-\tfrac{4}{3} p_n)^2 N].$$

Nei and Gojobori (1985) have shown that these equations give an accurate estimate of *dn/ds* ratio when the transition (T↔C and A↔G) and transversion (T,C↔A,G) ratio (*R*)

is not biased *i.e.* R=1. However, when the bias is large, *ds* tends to underestimate the true number of substitutions, as transitions at the third position in the codon are more likely to be synonymous than transversions. Therefore, the S is underestimated and N is overestimated (Kondo *et al.* 1993; Kumar *et al.* 1993). Li *et al.* (1985) developed a method in order to incorporate the transition/ transversion bias.

Li *et al.* method

Li *et al.* (1985) method allocates each site in the codon to 0-fold (non-degenerate), 2-fold (twofold degenerate) and 4-fold (fourfold degenerate) degenerate categories. The 0-fold includes all sites at which all changes are non-synonymous while the 2-fold includes those sites at which one out of three changes is synonymous and in the 4-fold those at which all changes are synonymous (Li *et al.* 1985). The observed transitional and transversional differences between codons are then partitioned into those occurring at 0-fold, 2-fold and 4-fold degenerate sites and changes between codons are assumed to occur with different probabilities (Li *et al.* 1985). At the twofold degenerate sites, transversions lead to non-synonymous changes whereas transitions lead to synonymous changes. In nuclear DNA there are two exceptions to this rule: the first positions of CGR and AGR (all code for Arg; R denotes any of A, T, G, C) and the last positions of ATH (all code for Ile; H denotes any nucleotide except G). In order to overcome this, Li *et al.* (1985) suggested an *ad hoc* adjustment where all synonymous changes in these cases are classified as transitions and all non-synonymous changes as transversions. The equations for *dn* and *ds* suggested by Li *et al.* (1985) are:

$$dn = \frac{3[L_2A_2 + L_4(A_4+B_4)]}{(L_2 + 3L_4)} \quad \text{and} \quad ds = \frac{3[L_2B_2 + L_0(A_0+B_0)]}{(2L_2 + 3L_0)}$$

where L_0, L_2 and L_4 are the number of 0-fold, 2-fold and 4-fold degenerate sites respectively and A_i and B_i are the number of transitional and transversional substitutions per *i*th type site and are given by the equations:

$$A_i=(1/2)\ln(a_i) - (1/4)\ln(b_i) \quad \text{and} \quad B_i=(1/2)\ln(b_i)$$

where $a_i=1/(1-2P_i-Q_i)$ and $b_i=1/(1-2Q_i)$

and P_i and Q_i are the number of observed transitional and transversional differences respectively, at the *i*th degenerate site divided by the L*i*, which is the total number of *i*-

fold degenerate sites, where i=0, 2 or 4. The equations for Ai and Bi are based on Kimura (1980) two-parameter method to estimate the mean of transitional and transversional substitutions rate per site, based on the observed proportions of transitional and transversional differences.

The program MEGA version 2.1 (Kumar *et al.* 2001) was used in order to estimate the *S*, *N*, *ds*, *dn*, *R*, ω values and their variances, according to the two approximate methods included in this study, Nei and Gojobori (1985) and Li *et al.* (1985).

2.6.1.2 Maximum likelihood methods

Previous studies in molecular evolution have suggested that the approximate methods can lead to an overestimation of ω value through codon usage bias (*e.g.* Goldman and Yang 1994; Nielsen and Yang 1998; Yang and Bielawski 2000; Yang and Nielsen 2000). The approximate methods assume that the rate of codon usage in the sequence is uniform. However, Yang and Nielsen (1998, 2000) have shown that codon-usage bias may significantly overestimates *S*, and underestimate *ds* which results in an overestimation of ω. The maximum likelihood methods count for both the transition/transversion bias and the codon-usage bias, therefore the ω value estimates are suggested to be a significantly improved fit to the real data (Goldman and Yang 1994; Nielsen and Yang 1998).

The unit of molecular evolution in the maximum likelihood methods is the codon and the following assumptions are made (Goldman and Yang 1994; Nielsen and Yang 1998): a) molecular evolution acts at the single site level, b) stop codons are not included as they are considered not to be involved in evolution, c) mutations at the three codon positions occur independently *i.e.* only single-nucleotide substitutions are permitted to occur simultaneously (with the exception of Kosakovsky *et al.* method) and d) the DNA sequences are presumed to be protein-coding. A Markov process is used in order to describe substitutions between the possible 61 codons (Goldman and Yang 1994). The substitution rate from codon i to codon j is defined (Goldman and Yang 1994; Nielsen and Yang 1998) as:

$$q_{ij}= \begin{cases} 0, & \text{if } i \text{ and } j \text{ differ at more than one position} \\ \pi_j, & \text{if } i \text{ and } j \text{ differ by a synonymous transversion} \\ \kappa\pi_j, & \text{if } i \text{ and } j \text{ differ by a synonymous transition} \\ \omega\pi_j, & \text{if } i \text{ and } j \text{ differ by a non-synonymous transversion} \end{cases}$$

$\omega\kappa\pi_j$, if i and j differ by a non-synonymous transition

where π is the equilibrium frequency of codon j, κ is the transition/transversion rate ratio and ω is *dn/ds* ratio. This substitution rate accounts for the three different factors observed in adaptive molecular evolution of protein-coding genes: a) the transition/transversion substitution rates are multiplied by κ which is the transition/ transversion rate ratio [estimated by Kimura (1980) method, $\kappa=2q$, where $q=R$] as transitions are known to occur more often than transversions (Goldman and Yang 1994), b) the codon-usage is taken into account by multiplying the substitutions rates by the equilibrium frequency of the given codon π_j and c) the possibility that the rates of non-synonymous and synonymous substitutions are unequal is taken into account by multiplying the substitution rate to the ω ratio. The parameter π_j is estimated using the averages of the observed frequencies.

The log-likelihood equation is:

$$l(t,\kappa,\omega,\pi)= \sum_{h=1}^{n} \log\{f(x_h)\}$$

and $$f(x_h)= \pi_{x1h}\, p_{x1h,x2h}\,(t_1+t_2)$$

Yang-Nielsen method

The Yang and Nielsen (2000) method is a full likelihood approach which models variation in non-synonymous and synonymous sites according to a pre-defined model of distributions. There are nine different substitution rates models (Yang 1997) however, for the purpose of this investigation I will focus on the HKY85 model which is the model recommended for the present data obtained through the model selection procedure of the PAML program version 3.14 (Yang 1997). The matrix for this model was:

$$Q= \begin{pmatrix} . & \kappa\pi_C & \pi_A & \pi_G \\ \kappa\pi_T & . & \pi_A & \pi_G \\ \pi_T & \pi_C & . & \kappa\pi_G \\ \pi_T & \pi_C & \kappa\pi_A & . \end{pmatrix}$$

Adaptive selection can be detected by using a series of models which constraint ω to fixed values or a distribution. The models used in the present study were those recommended by Yang *et al.* (2000):

***M0** (one ratio)*	This model assumes one ω among all sites (Goldman and Yang 1994).
***M1** (neutral)*	This model assumes that ω either equals one or zero to represent the neutral theory of evolution (Nielsen and Yang 1998).
***M2** (selection)*	This model is the same as M1 but there is an additional class of ω which is free to take any value (Nielsen and Yang 1998)
***M3** (discrete)*	This model uses an unconstrained discrete distribution of ω with a set number of classes: ω=1, ω<1 and ω>1
***M7** (beta)*	This model assumes a distribution of ω values between 0 and 1 among the sites
***M8** (beta and ω)*	This model is the same as M7 except ω is a free parameter

A maximum likelihood logarithm was estimated by each model and a likelihood ratio test (LRT) was performed in order to determine statistically which model fits the data significantly. For example given two models A an B, A with p parameters and log likelihood of l_1 and B with q parameters and log likelihood l_2, then $2(l_1-l_2)$ can be compared to a χ^2 distribution with degrees of freedom p-q (Stuart *et al.* 1999). If $2(l_1-l_2)$ is sufficiently large then l_2 is rejected. In addition, each model estimates the values of *S*, *N*, *ds*, *dn* and ω according to maximum likelihood equations (see Yang 1998).

The Yang-Nielsen method was performed by PAML program version 3.14 (Yang 1997) in the implemented program called *codeml*, the phylogenetic tree required by the imput file was generated through MEGA version 2.1 (Kumar *et al.* 2001). In order to reduce the computational time required by *codeml* two simplifications were taken (Yang 1998): a) the frequency of codon *j*, π_j was calculated from the observed base frequencies at the three codon positions and b) *M0* model was executed first with the supplied phylogenetic tree then the phylogenetic tree inferred by this model was used for all the other models.

Kosakovsky *et al.* method

This maximum likelihood method is similar to the *M3* model of Yang-Nielsen method with the following modifications: a) the synonymous and non-synonymous substitutions are allowed to vary from site to site as a lack of this provision may lead to a high rate of false-positive results (Kosakovsky and Muse 2004), b) a more general model of nucleotide substitution rates is used, rather than estimating the model from the data, c) the Muse-Gaut model for the frequency components (κ,ω,π) is used (Kosakovsky and Muse 2004) and d) Bayes factors (inferring probabilities for the phylogeny) are used.

2.6.1.3 Radical and conservative substitution analysis

Amino acids can be classified according to their physico-chemical properties into three different groups: a) charge, b) polarity and c) volume. Substitutions within groups are referred to as conservative while substitutions between groups as radical. Previous studies have suggested that, for genes exhibiting high ω ratios and strong evidence of positive selection, the estimation of the radical/ conservative (*dr/dc*) ratio may assess the nature of selection (Hughes *et al.* 1990; Zhang 2000; Zhang *et al.* 2002; Smith 2003). Zhang *et al.* (2002) showed that substitutions in pancreatic ribonuclease gene resulted in an elevated (significantly different than one) *dr/dc* ratio with the charge classification and not with polarity or volume classification, supporting the hypothesis that selection has acted in order to increase enzymatic activity in low pH. This was further supported by the finding that all charge-altering substitutions increased the negative charge of the gene (Zhang *et al.* 2002).

For the purpose of the present study, amino acids were classified according to Zhang (2000) method:

Charge	**Polarity**	**Polarity and Volume**
Positive: *R, H, K*	**Polar:** *R, N, D, C, Q, E, G*	**Special:** *C*
Negative: *D, E*	*H, K, S, T, Y*	**Polar and small:** *N, D, Q, E*
Neutral: *A, N, C, Q, V,*	**Non-polar:** *A, I, L, M, F, P*	**Polar and large:** *R, H, K*
G, I, L, M, F, P, S, T, W, Y	*W, V*	**Non-polar and small:** *I, L, M, V*
		Non-polar and large: *F, W, Y*
		Neutral and small: *A, G, P, S, T*

The computation of radical and conservative substitutions for each of the classifications follows the substitution method of Nei and Gojobori (1986) described

previously and the *C*, *R*, *dc*, *dr* and *dr/dc* are derived by an analogy of that method. The difference is that the transition/ transversion ratio is taken into account:

$$c_i = \sum_{j=1}^{3} c_{ij}$$

where c_{ij} is the expected number of conservative non-synonymous changes when a nucleotide change occurs at the j_{th} (j= 1, 2 or 3) position of the i_{th} codon and substitutions follow Kimura model (1980). The number of radical (r_i) non-synonymous sites for codon *i* is computed similarly. Therefore, the number of non-synonymous sites for the *i*th codon is given by $c_i + r_i = n_i$ where parameters c_i, r_i are multiplied by q (transition/transversion ratio in Kimura model) to account for transition/transversion bias. The HON-NEW program (Zhang 2000) was used for estimating conservative and radical nucleotide substitutions.

2.6.1.4 Categorization of MHC alleles on a functional basis

The molecular structure of HLA-DR has identified subregions (referred to as pockets) in the *DR* binding groove which exert major influence on the binding, presentation and recognition by T-cells (Stern *et al.* 1994; Ou *et al.* 1998). Among these subregions, pocket 4 amino acid residues β70, β71 and β74 are located in the *DRB* chain and are thought to play a significant role in determining T-cell recognition of the peptide-*DR* complex (Olson *et al.* 1994; Stern *et al.* 1994; Ou *et al.* 1996).

The alleles of the *DRB* locus of HLA class II molecules have been grouped into seven different functional categories according to physicochemical polymorphisms of the amino acid residues β70, β71 and β74 of pocket 4 in the peptide binding region and their influences on T-cell recognition (Ou *et al.* 1997; Ou *et al.* 1998). Previous studies have shown that the charge of the residues of those categories influence the selectivity for peptide binding (Hammer *et al.* 1995; Wecherpfennig and Strominger 1995; Wecherpfennig *et al.* 1995; Ou *et al.* 1996). Ou *et al.* (1998) suggested that these seven categories can be combined into four groups based on the charges of the β70, β71 and β74 residues: a positively charged group (+), a negatively charged group (-), a di-charged group (+/-) and a neutral group (n).

According to Ou *et al.* (1998) method, the charge of pocket 4 is estimated according to the sum of β70, β71 and β74 residues charge. For example, the

*DRB1**1117 allele exhibits the residues RRE in positions 70, 71 and 74 respectively. Arginine (R) is positively charged and glutamic acid (E) is negatively charged and this allele is classified in the di-charged functional group. When a charged amino acid is present among non-polar and/or neutral amino acids, then the allele is classified according to the charged amino acid present. The charge of the amino acids was determined according to the following categorization (Ou *et al.* 1998):

Neutral	**Positively charged**	**Negatively charged**
F, M, W, I, V, L, A, P, C, N, Q, T, Y, S, G	*H, K, R*	*D, E*

2.6.1.5 Estimating gamma shape parameter α

The gamma distribution is used in order to model substitution rate variation among amino acid residues in protein-coding genes (Jin and Nei 1990; Yang 1994; Miyamoto and Fich 1996). Under this model the variation of substitution rate λ is estimated by:

$$\Phi(\lambda) = \frac{\beta_a}{\Gamma_{(a)}} \lambda^{a-1} e^{-\beta\lambda}$$

where the shape parameter α is important because it describes the degree of variation and $1/\sqrt{\alpha}$ is the coefficient of a variation λ (Tamura and Nei 1993; Yang 1993; Gu and Zhang 1997). Therefore, the larger the value of α, the weaker the rate of variation λ and a when $\alpha=\infty$ indicates a uniform rate among sites. Furthermore, although the biological meaning of the parameter is still not well defined, α value has been used as a further support for selection inferred by substitution analyses (Gu and Zhang 1997). The ML approach for estimating α was developed by Gu and Zhang (1997) and the log likelihood function is:

$$\ln L = \sum_{i=1}^{n} \ln f(k_i)$$

where N is the total number of sites and k_i is the estimate of the expected number of substitutions at site i. The gamma parameter was estimated using GZ-GAMMA program (Gu and Zhang 1997). The input file contains the tree topology of the sequences which is used to infer the ancestral sequence for each node by a Bayesian approach (Zhang and Nei 1997). The expected number of substitutions per site is estimated by the Jukes-Cantor model (see 2.6.1). The estimate of α is obtained by the maximum likelihood approach described above. In the present investigation the main use of the inferred α value was to account for rate variation among sites in the phylogenetic analyses. This is achieved by re-computing the inferred phylogenetic trees including the gamma distribution for the rate variation among sites (Gu and Zhang 1997).

2.6.2 Phylogenetic analysis

Phylogenetic trees showing genetic distance were constructed using the program PAUP version 3.1 (Swofford 1991). This program enables the construction of trees taking into account the observed nucleotide frequency, substitution rate and transition/transversion ratio, based on Nei genetic distance estimates, maximum parsimony and maximum likelihood methods. The bootstrap analysis (Felsenstein 1988) is used in order to statistically assign support to hypotheses of phylogenetic relationship among sequences estimated by the methods mentioned above. This method involves sampling the original data set with replacement in order to construct a series of bootstrap replicates of the same size as the original data set (Swofford 1991). A majority-rule consensus tree is constructed for all of the bootstrap trees and the bootstrap values indicate the number of times a given group has appeared in the bootstrap reconstruction (number is given as a percentage; see Chapter 3: 3.2.2).

2.6.2.1 Neighbor-Joining method

Nei (1991) demonstrated by computer simulations that the Neighbor-joining method (Saitou and Nei 1987) is one of the most efficient distance methods. In this method, the topology showing the smallest value of the sum S of all branches is chosen as an estimate of the correct tree, where $S= 2m-3$ for a bifurcating tree with m operational taxonomic units (OTU). However, the S value is not calculated for all the

possible topologies, but the examination of different topologies is included in the Neighbor-Joining algorithm so that only one final tree is produced (Saito and Nei 1987).

2.6.2.2 Maximum parsimony method

Trees constructed by maximum parsimony account the nucleotide sites at which there are at least two different kinds of nucleotides present, each represented at least twice (referred to as parsimony-informative sites; Kumar *et al.* 1993). Three OTUs are chosen initially and a tree is constructed containing only those. The length of the initial tree is the largest among all possible 3-OTUs determined by the data set *i.e.* this tree exhibits the highest number of nucleotide differences. Using this initial tree, the minimum number of substitutions that are required to explain the nucleotide differences among OTUs is counted. The sum of this number over all parsimony informative sites is the number of steps. This process is repeated and the tree of three OTUs which shows the largest value is chosen as the initial core tree. The remaining OTUs are fitted to the core tree and the tree which exhibits the smallest number of steps is chosen as the final tree.

2.6.2.3 Maximum likelihood method

Felsenstein (1981) introduced a maximum likelihood approach where trees are constructed based on their maximized probability of observing the data. The maximum algorithm is used for searching between different tree topologies and within topologies using a combination of the 'pruning' and 'pulley' principle (Felsenstein 1981). Pruning involves removing one pair of OTUs at a time and estimates the likelihood of the remaining OTUs while pulley involves altering the branch lengths since only relative differences are considered and therefore the lengths of two branches with a common node can be varied in tandem (Felsenstein 1981). The tree with the largest log likelihood of observing the given alignment under the chosen evolutionary model is chosen as the final tree. Log likelihood calculation takes into account branching order and branch lengths.

2.6.2.4 SLAC method

The Single Likelihood Ancestor Counting method was developed by Kosakovsky and Muse (2004). Given an estimate of the phylogeny and the codon-based substitution model of a set of sequences, this method estimates the number of

changes which have occurred along the phylogeny by using the joint reconstruction method of Pupko *et al.* (2000). The final tree is a phylogenetic reconstruction of codon substitutions by incorporating synonymous and non-synonymous substitution differences into the ancestral state reconstruction method. The ancestral reconstruction uses extant sequences and the phylogenetic relationships among them in order to infer the most possible ancestral sequences from which they have been derived. Maximum likelihood is used in order to estimate the probability of inferred phylogeny. The program Hyphy (Kosakovsky and Muse 2004) was used in order to construct the SLAC tree phylogeny, using the Neighbor-Joining tree constructed by PAUP version 3.1 (Swofford 1991) and the codon substitution model inferred by PAML version 3.14 (Yang 1997).

2.6.2.5 Median joining network method

The median joining method begins with a minimum spanning network, which is a tree for a set of sequence types connected without assuming any phylogenetic relationships (in the form of cycles) or inferring additional ancestral nodes, calculated by the Kruskal algorithm (Bandelt *et al.* 1999). The sequences are combined in a single network. The next step adds a few consensus sequences (median vectors) of three mutually close sequences determined by parsimony. The median vectors are interpreted as extant un-sample sequences or extinct ancestral sequences. The approach of Bandelt *et al.* (1999) uses the median vectors which have a good chance of appearing as branching nodes in a Maximum parsimony tree, determined by the sequences of the network under processing. The procedure is repeated until no new sequences can be added. The final tree is a multitude of plausible trees and displays alternative potential evolutionary paths in the form of cycles. The program NETWORK version 2.0 (Rohl 1997) was used to generate median joining networks.

2.6.3 Recombination analysis

2.6.3.1 Split decomposition analysis

The approach of split decomposition analysis is somewhat between phylogenetic trees inferred by distance methods or maximum parsimony or maximum likelihood and the median joining network method. Split decomposition represents all conflicting phylogenetic signals in the form of alternative network pathways. For example,

evolutionary relationships are represented by a phylogenetic tree whose lineages are labeled by a given set of OTUs (or taxa). The edge of those lineages represents a split. The split decomposition algorithm exhibits less restricted systems of splits allowing a more general graph which in turn represents a weakly compatible split system. This method depicts parallel edges between sequences on the basis that there are conflicting phylogenetic signals in the data. Split decomposition method is used in order to determine whether recombination has affected observed phylogenetic relationships, as recombination in DNA sequences produces networks of sequences rather than strictly bifurcating evolutionary trees (Fitch 1997; Huson 1998; Holmes *et al.* 1999). The program SplitsTree 4-beta5 version (Huson and Bryant 2004) was used in order to estimate the phylogenetic tree inferred by the split decomposition method.

2.6.3.2 Likelihood analysis of DNA recombination

The likelihood analysis is used in combination to the split decomposition analysis in order to map individual recombination events and to provide the statistical support for those events. The likelihood analysis of DNA recombination estimates the branch lengths of a tree topology of the putative recombinant and two parental sequences (Holmes *et al.* 1999). A break-point (κ) is chosen along the branch lengths and by moving the κ point across the alignment different topologies are produced. Therefore, by moving the breakpoint in the three sequences (with three free parameters) a different tree is produced with six free parameters. By moving κ the maximum likelihood position of the recombination breakpoint is determined. A Monte Carlo stimulation (500 replications) is used to assess the significance of the fit of κ to the data. The program LARD version 2.2 (Holmes *et al.* 1999) is used for maximum likelihood analysis of recombination.

2.6.4 Population genetic analysis

2.6.4.1 Hardy-Weinberg equilibrium

According to the Hardy-Weinberg equilibrium (HWE), the genotype frequencies in a population of random mating diploid individuals which is free of migration, mutation, genetic drift and selection will be equal to: $p^2 + 2pq + q^2 = 1$ (see review by Gillespie 1998). In the case of a population exhibiting allelic frequencies significantly different from the HWE values, non-random breeding (inbreeding), physical linkage,

non-random sampling of panmictic individuals from a large population or the Wahlund effect may be invoked. Inbreeding and population subdivision will cause an increase in the number of homozygotes within the population while outbreeding will increase a number of heterozygotes within the population. The observed and expected according to HWE levels of heterozygosity should be not significantly different. The departure of genotype frequencies from the predictions of HWE may imply the effect of selection (Gillespie 1998).

Genetic diversity as a measure of individual variation within a population reflects the number of different types in the population, taking into account their frequencies (Gregorious 1987). Levels of genetic diversity for each population were examined using the program ARLEQUIN version 2.000 (Schneider *et al.* 2000). The number of unique alleles was calculated for each locus and over all loci and the observed heterozygosity (H_O) and expected heterozygosity (H_E) levels at each locus were tested for deviation from HWE using Fisher's exact test, implemented in the ARLEQUIN program. The probability of Fisher's exact test is tested using the Markov chain method with 1000 dememorizations, 100 batches and 1000 iterations. The equation for the unbiased expected heterozygosity level at a given locus in a given population is:

$$H_E = \frac{2n(1-\sum P_i^2)}{(2n-1)}$$

where p_i is the frequency of each of the alleles at a locus and n is the number of sampled individuals (Nei 1987).

2.6.4.2 *F* statistics

Wright (1951) devised methods to test the genetic population structure in terms of three F-statistics: F_{IS}, F_{IT} and F_{ST}. Theses indexes are used in order to distinguish between three possible levels of inbreeding and to quantify population substructure.

F_{IS}

Wright's F_{IS} is the correlation between homologous alleles among individuals that are part of a local population (Avise 1994). Non-random mating (*e.g.* mating

between related individuals) is one cause of the reduction in the heterozygosity of a n individual within a population. The program FSTAT v. 2.9.3 (Goudet 2001) was used in order to calculate F_{IS} and the equation is:

$$F_{is} = \frac{H_s - H_i}{H_s}$$

The degree of non-random mating (inbreeding) within a population was assessed by comparing the observed and expected heterozygosity levels where, H_i is equal to the observed heterozygosity of an individual, estimated as the mean frequency of heterozygotes averaged over all subpopulations and H_s is the expected heterozygosity of an individual in a subpopulation, calculated separately for each subpopulations and then averaged. F_{IS} was initially developed for a locus with only two alleles, therefore, Weir and Cockerham's (1984) estimator (*f*) of *Fis* was calculated in order to develop a method for many loci and multi allelic data:

$$f = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

Where σ_b^2 is the between individual variance component and σ_w^2 the within individual component. The estimator (*f*) of *Fis* was calculated using FSTAT v 2.9.3 (Goudet 2001).

***F*_{*st*}**

This index assumes an infinite allele mutation model in which every new mutation creates a novel allele. F_{st} is based on the variance in allele frequencies to determine the level of population subdivision caused by a reduction in heterozygosity (Avise 1994). This parameter varies from 0 (no differentiation) to 1 (complete differentiation) and is estimated using pairwise population comparisons for the proportion of variance that accounts for between rather than within population differences (Nei 1973). F_{st} is defined as:

$$F_{st} = \frac{(H_t - H_s)}{H_t}$$

Where H_t and H_s are proportional to the expected heterozygosity of an individual in the whole and subpopulation respectively. When calculating H_t it is assumed that all the samples were selected from one homogeneous randomly mating population. H_s is first calculated for each subpopulation and then averaged together. F_{st} is dependent on the assumptions that populations are maintained under the same conditions, they are derived from a common ancestor and that gene frequencies are at a state of equilibrium (Tufto *et al.* 1996). Mutational relationships among alleles are not considered. Statistical significance was tested by 10,000 permutations of the data. The estimator (θ) of F_{st} of Weir and Cockerman (1984) is required in order to estimate F_{st} for a locus with more than two alleles:

$$\theta = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2 + \sigma_w^2}$$

where σ_a^2 represents the among sample variance, σ_b^2 is the between individuals variance and σ_w^2 is the within individuals variance. Values of θ were calculated by the program ARLEQUIN version 2.0 (Schneider 2000) and FSTAT v. 2.9.3 (Goudet 2001) and a permutation based approach was used by both programs in order to test the values of θ. If differentiation between putative populations is nonexistent, then members from each population should be able to split and randomly assigned to new groups. This should generate values of F_{st} comparable to the original.

2.6.4.3 Relatedness

Relatedness (*Rel*) between individuals was estimated by using the measure of Hamilton (1971) with the estimator of Queller and Goodnight (1989). The program FSTAT v.2.9.3 (Goudet 2001) was used: $Rel = 2\ F_{st}/(1+F_{it})$. The F_{it} index is calculated as:

$$F_{it} = \frac{H_t - H_i}{H_t} \quad \text{and} \quad F_{it} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_w^2}$$

where H_i is the observed heterozygosity of an individual estimated as the mean frequency of heterozygotes averaged over the whole population while H_t is the expected heterozygosity of an individual in the whole population. However, Pamilo (1984) demonstrated that inbreeding can affect the estimation of *Rel* and in order to account for the inbreeding bias the corrected relatedness (*Rel'*) is estimated by:

$$Rel' = \frac{\left(\frac{rel - 2F_{it}}{1 + F_{it}}\right)}{\left(\frac{1 - 2F_{it}}{1 + F_{it}}\right)}$$

The program FSTAT v. 2.9.3 (Goudet 2001)was used to estimate *Rel* and *Rel'*.

2.6.4.4 Allelic richness

The number of alleles found within a given population sample is highly dependent on the sample size of the population. Allelic richness (R) for each population was measured to insure that the observed number of alleles found within a population is independent of the sample size of the population. The concept is to estimate the expected number of alleles in sub-sample of $2n$ genes, given that $2N$ genes have been sampled ($N \geq n$). The smallest number of individuals typed for a locus in a sample is fixed. The values of n and R are calculated with the program FSTAT version 2.9.3 (Goudet 2001). The equation is:

$$R_s = \sum_{i=1}^{n_i} \left[1 - \frac{\left(\frac{2N - N_i}{2n}\right)}{\left(\frac{2N}{2n}\right)} \right]$$

where N_i is the number of alleles of type i among the $2N$ genes. The terms under the sum correspond to the probability of sampling allele I at least once in a $2n$ sample size. When the frequency of allele i is very high then the ratio is undefined and the probability of sampling the alleles is set to 1.

2.6.4.5 Gene diversity

Nei (1987) defined a measure of gene diversity (H_s) as the probability that two randomly chosen haplotypes are different in the sample and is estimated as:

$$H_s = \frac{n}{n-1}\left(1-\sum_{i=1}^{k} p_i^2\right)$$

Where n is the number of gene copies in the sample, k is the number of haplotypes and pi is the probability of the i^{th} haplotype. Gene diversity was estimated by the program FSTAT v. 2.9.3 (Goudet 2001).

2.6.4.6 Neutrality analysis

The program ARLEQUIN version 2.000 (Schneider 2000) was used in order to perform the Ewens-Watterson homozygosity test. Watterson (1978) has shown that the distribution of neutral haplotype frequencies can be presented as the sum of haplotype frequencies (F), which is equivalent to the expected homozygosity. The allele frequency distribution (whether it is even or not) is determined by the observed homozygosity statistic, which is the observed allele frequencies rather than the observed homozygote frequencies, and it is estimated by the equation:

$$\hat{F} = \sum_{i=1}^{k} \frac{n_i^2}{n^2}$$

where, $\frac{n_i^2}{n^2}$ is the probability of the frequency of the i^{th} allele. Therefore, for a given population, in the case of a low value of the homozygosity statistic the allele frequencies are even distributed while in the case of a high value of the statistic the allele frequencies are un-ever distributed. The comparison between the observed and

expected under neutrality value of F (estimated by the algorithm of Steward 1977) provides a test of neutrality. The Ewens-Watterson-Slatkin exact test which is implemented in ARLEQUIN program estimates the probability of obtaining a random sample having a probability of smaller or equal to the observed sample (Schneider 2000).

2.6.5 General statistics used

The MINITAB™ version 9 was used in order to estimate means and standard deviations for the data. This program was also used in order to perform t-test, Spearman correlation test and goodness of fit tests. When multiple statistical tests were performed, the Bonferonni correction was used in order to account for chance capitalization and it is estimated by: α/n where α is the chosen significance level and n is the number of statistical comparisons. ARLEQUIN version 2.000 (Schneider 2000) was used in order to perform the Mantel matrix correlation test, which assesses the significance of the correlation between two matrices by a permutation procedure.

CHAPTER 3:

EVOLUTION OF MHC *DQB1* EXON-2 IN CETACEANS

3.1 INTRODUCTION AND OBJECTIVES

The human impact on the marine environment has caused the extinction of several marine mammal species, for example the Steller's sea cow (*H. gigas*), Japanese sea lion (*Z. c. japonicus*) and Caribbean monk seal (*M. tropicalis*) and the depletion of many other cetacean species (see review by Martin and Reeves 2002). It was not until late 1960's that conservation efforts and regulations where applied at a multi-species level (see review by Taylor 2002). Prior to that, all efforts where on specific species and populations. For example, in the 1920's, the whaling industry had begun to place limits on oil production and gave full protection to the right and gray whales, allowing fleets to redirect their attention on other species like humpback, blue, fin and sei whales (see review by Reeves 2002). The Marine Mammal Protection Act was erected by the American congress in 1972 and was one of the first laws passed as a national commitment to conservation. The goal was to achieve and maintain "optimum sustainable populations" of marine mammals and conserve genetic diversity, as this is the parameter that will allow a given species to evolve and adapt to a changing environment.

In order for conservation to be applied efficiently there are many different aspects of the life history of a target species, which are taken into consideration, such as reproduction, mortality and natality, survivorship, foraging behavior and age distribution (Taylor 2002). Cetaceans are still at a great risk and this is mainly due to the fact that they are slow reproducing, long lived and large bodied species, which make them especially prone to extinction (Hoelzel *et al.* 2002).

Mass strandings and large-scale mortalities caused by pathogens have been recorded in recent years and have shown that pathogens are capable of threatening and depleting cetacean populations (see Chapter 1). Van Bressem *et al.* (1999) suggested that mass mortalities by pathogens are the most important factor in determining long-

term population sizes in the absence of human exploitation. The genetic study of the MHC can be very informative not only at an evolutionary genetics level but also at a conservation level. For example, O'Brien *et al.* (1985), Yunki and O'Brien (1990) and Hughes (1991) have suggested that MHC variation should be taken into account in conservation and management programs, as MHC variation deficiency may increase the risk of extinction of small isolated populations.

3.1.1 MHC evolution in Marine Mammals

Trowsdale *et al.* (1989) used human class I and class II probes to investigate MHC variation in two *Mysticeti* species, the fin (n=9; *Balaenoptera physalus*) and sei whale (n=5; *B. borealis*). Restriction fragment length polymorphism (RFLP) analysis was performed and it was suggested that MHC in whales does not differ radically from human. However, this study also suggested low levels of polymorphism compared to human and mouse, an observation which was hypothetically attributed to weak selection pressures due to less body contact in cetaceans than in terrestrial mammals.

Slade (1992) investigated MHC polymorphism in two different populations of elephant seals (*Mirounga leonina*) using RFLP analysis and showed low levels of MHC polymorphism. This observation was also argued to be the result of diminished exposure to pathogenic selection pressure (Slade 1992). However, a study on bottlenose dolphin (*Tursiops truncatus*) by Romano *et al.* (1992) investigated the expression of class II molecules on lymphocytes and they showed that 100% of the peripheral T-cells exhibit pronounced expression of class II molecules, unlike the majority of land mammals (22%; Romano *et al.* 1992) and comparable to ungulates and carnivores (Romano *et al.* 1992).

Murray *et al.* (1995) investigated the genetic variation of MHC *DQB1* locus exon-2 in seven populations in beluga whale (n=233; *Delphinapterus leucas*). They used PCR and SSCP analyses in order to sequence the Peptide Binding Region (PBR). The nucleotide substitution analysis demonstrated that there was a higher ratio of non-synonymous substitutions in PBR sites. In addition, Murray *et al.* (1995) suggested that, although the amount of allelic variation was low (n=5) in beluga compared to terrestrial mammals, the substitution analysis indicated that variation in cetaceans may be attributed to positive selection.

Furthermore, a study by Hoelzel *et al.* (1999) in MHC *DQB1* exon-2 variation in southern elephant seal (n=109; *Mirounga leonina*), northern elephant seal (n=69; *Mirounga angustirostris*), Antarctic fur seal (n= 13; *Arctocephalus gazella*) and New Zealand fur seal (n=19; *Arctocephalus fosteri*) showed that the phylogenetic analysis of these species was consistent to the trans-species evolution pattern observed in terrestrial mammals (Klein 1986). In addition, the nucleotide substitution pattern in exon-2 suggested that *DQB1* variation is maintained by positive selection, as higher non-synonymous rates were observed (Hoelzel *et al.* 1999). They suggested that the pattern of evolution of MHC in the four species of pinnipeds is comparable to that in terrestrial mammals. This study also used PCR, SSCP and DNA sequencing analyses in order to assess genetic variation. Flores-Ramirez *et al.* (2000) investigated the PBR sites of class I MHC molecules in gray whale (*Eschrichtius robustus*) and showed that nucleotide variation in three class I loci exhibited substitution patterns similar to that attributed to selection in terrestrial mammals.

Hayashi *et al.* (2003) studied *DQB1* exon-2 variation (PCR-SSCP approach) in five species of *Mysticeti* (n=9) and eleven species of *Odontoceti* (n=22) and they suggested that although cetacean *DQB1* alleles exhibit low levels of trans-species evolution, the evolution pattern is consistent to that observed in terrestrial mammals. In addition, they showed that non-synonymous substitution rates in exon-2 are comparable to those in terrestrial mammals.

3.1.2 Objectives

The divergence in mammals occurred in an early stage allowing different selection pressures to act and affect evolution pattern in cetaceans (see Chapter 1; Thewissen 1998; Thewissen and Hussain 1998; Heyning and Lento 2002). Previous studies in cetaceans have suggested low levels of variation in MHC, an observation which may be attributed to the fact that a single species was investigated (such as Beluga whale in Murray *et al.* 1995) or to low levels of resolution of the method used (such as RFLP in Trowsdale *et al.* 1989). For the purpose of the present study, a thorough sequence analysis using the PCR-SSCP method is conducted, using an extensive sample of cetacean species. The main aim is:

I) **To evaluate the evolutionary pattern and radiation of *DQB1* locus in the order *Cetacea*.** This will be assessed through phylogenetic and

nucleotide analysis. The purpose of the phylogenetic analysis is to determine the intraspecific and interspecific phylogeny of the identified alleles, as overdominant and/or frequency-dependent selection will result in polytomy and a trans-specific evolutionary pattern. Nucleotide analysis will assess the nature and pattern of amino acid substitutions in the i) full, ii) PBR and iii) non-PBR sites of exon-2 sequence. According to the expectations of balancing selection, sites involved in antigen presentation will exhibit higher rates of non-synonymous substitutions. The main question is whether MHC genes in cetaceans are under homologous selection pressure and thus have gone through a similar evolutionary path as in terrestrial mammals.

3.2 RESULTS

For the purpose of this study, *DQB1* variation was assessed through sequence and phylogenetic analysis. The nature, pattern and rate of amino acid substitutions were determined through sequence analysis, as a higher rate of non-synonymous to synonymous and radical to conservative substitutions is observed in MHC *DQB1* loci in other mammalian species (Chapter 1), as the result of positive selection. Putative PBR and non-PBR sites are taken into consideration.

3.2.1 Sequence analysis

A total of 573 individuals from 18 cetacean species was surveyed in order to isolate and determine *DQB1* exon-2 sequences. The oligonucleotide primers *DQB1*-PBR, developed by Tsuji *et al.* (1992), are based on conserved regions flanking the PBR of exon-2. All alleles were identified and distinguished through SSCP and confirmed by sequencing. The comparison between each of the alleles to the complete sequence database in Genebank (BLAST; Zhang and Madden 1997) revealed sequence homology of up to 92% with published sequences of ungulate species. The highest homologies were observed between horse, swine, sheep and donkey *DQB1* exon2. Table 3.2.1 summarizes the samples analyzed. The absolute sequence difference between alleles was up to 11% within species and 12.3% among species.

Six alleles out of seven are shared between *T. truncatus* and *T. aduncus* and four alleles out of ten between *B. bonaerensis* and *B. acutorostrata*. MHC allele sharing

between closely related species has been observed both in marine and terrestrial mammals in previous investigations (*e.g.* Ellis *et al.* 1999; Hoelzel *et al.* 1999; Van Der Walt *et al.* 2001). Furthermore, only four alleles were observed in *B. acutorostrata*, compared to ten in *B. bonaerensis*, although the sample sizes were the same. In *O. orca*, *T. truncatus*, *B. bonaerensis* and *B. acutorostrata* species, samples are derived from the majority of identified populations determined by recent studies, representing a wide geographic range of populations. The population genetic analysis of each of the aforementioned species is presented in Chapter 5.

Although *DQB1* alleles were easily identified by SSCP, it has been showed that recombinant *DQB1* sequences can be produced by mismatch repair of heteroduplexes during cloning in *Escherichia coli*, an observation, which was previously thought to be

Table 3.2.1: The suborder and family name of all the species involved in this study. *N* is the number of individuals from each species and the number of alleles identified through SSCP and sequencing.

Suborder	Family	Species	N	No. of alleles
Odontoceti	*Delphinidae*	*Orcinus orca*	181	7
		Tursiops truncatus	90	7
		Tursiops aduncus	140	7
		Stenella frontalis	1	1
		Stenella coeruleoalba	1	1
		Delphinus delphis	1	1
		Lagenorchynus acutus	1	1
		Grampus griseus	1	1
		Globicephala melas	1	2
	Phocaenidae	*Phocoena phocoena*	2	3
	Ziphiidae	*Mesoplodon stejneri*	1	1
		Ziphius cavirostris	1	1
		Mesoplodon densirostris	1	1
	Physeteridae	*Physeter macrocephalus*	4	3
Mysticeti	*Balaenopteridae*	*Megaptera novaeangliae*	27	8
		Balaenoptera bonaerensis	60	10
		Balaenoptera acutorostrata	59	4
		Balaenoptera physalus	1	1

due to *in vitro* recombination and artifact generation by the PCR (Longeri *et al.* 2002). In order to overcome this and to confirm that only one locus was amplified, 24 clones of each individual were subsequently characterized by SSCP and putative variants eliminated (see chapter 2: 2.5.2). Sequencing revealed no more than two sequences in each individual.

3.2.1.1 Sequence alignment analysis

The consensus of all alleles identified in this investigation contains a continuous open reading frame for the entire 171bp sequence. Figures 3.2.1 and 3.2.2 show the nucleotide sequences and the translation for each of the cetacean alleles, note that polymorphic sites cluster in the β sheet and the α helix of the protein, and also around the sites which are inferred to be important for the selective binding and presentation of antigens, a feature which is in accordance with data for terrestrial mammals and HLA. This fragment corresponds to the 21st to 77th amino acid residues of the exon-2 HLA class II *DRB1* locus and demonstrates the highest homology to HLA *DQB1* among all HLA class II loci. Furthermore, a maximum of two different sequences was found in all individuals included in this study, which indicates that sequences are derived from the successful amplification of only one locus.

Site 37 (Figure 3.2.2) displays a high degree of polymorphism, even though it is not considered to be involved in antigen presentation based on previous studies of HLA *DQB1*. The Pocket 4 (P4) region in the PBR is comprised of β70, β71 and β74 amino acid residues and in HLA *DRB* studies have shown that differences in these sites have a profound effect on the actual topology and binding between the antigen and PBR (e.g. Olson *et al.* 1992; Stern *et al.* 1994; Fu *et al.* 1995). Subsequent presentation is affected to a great extent by non-synonymous substitutions at these particular residues (*e.g.* Wecherpfennig *et al.* 1995; Ou *et al.* 1998). The cetacean alleles exhibit high levels of variation in these sites where 88% of substitutions occurring in the putative Pocket 4 region in cetaceans are non-synonymous.

Figures 3.2.3 and 3.2.4 show the nucleotide and amino acid variation rate for single positions in exon-2 respectively. Both figures clearly illustrate that there is a higher degree of polymorphism in the PBR sites than in the non-PBR and also that variation in exon-2 is concentrated in the β-sheets and the α-helix of the protein.

Figure 3.2.1: *DQB1* exon-2 nucleotide sequence for all the species and alleles involved in this study. Only the sites exhibiting nucleotide variation are shown. Asterisk denotes codons important in antigen presentation and codons in bold font style polymorphic regions in HLA *DQB1* (Brown *et al.* 1993). P4, P7 and P9 residues indicated by red, blue and green respectively. Sequences suffixed with (*-c*) were provided by Hayashi *et al.* (2003) and with (*-j*) by Murray *et al.* (1995).

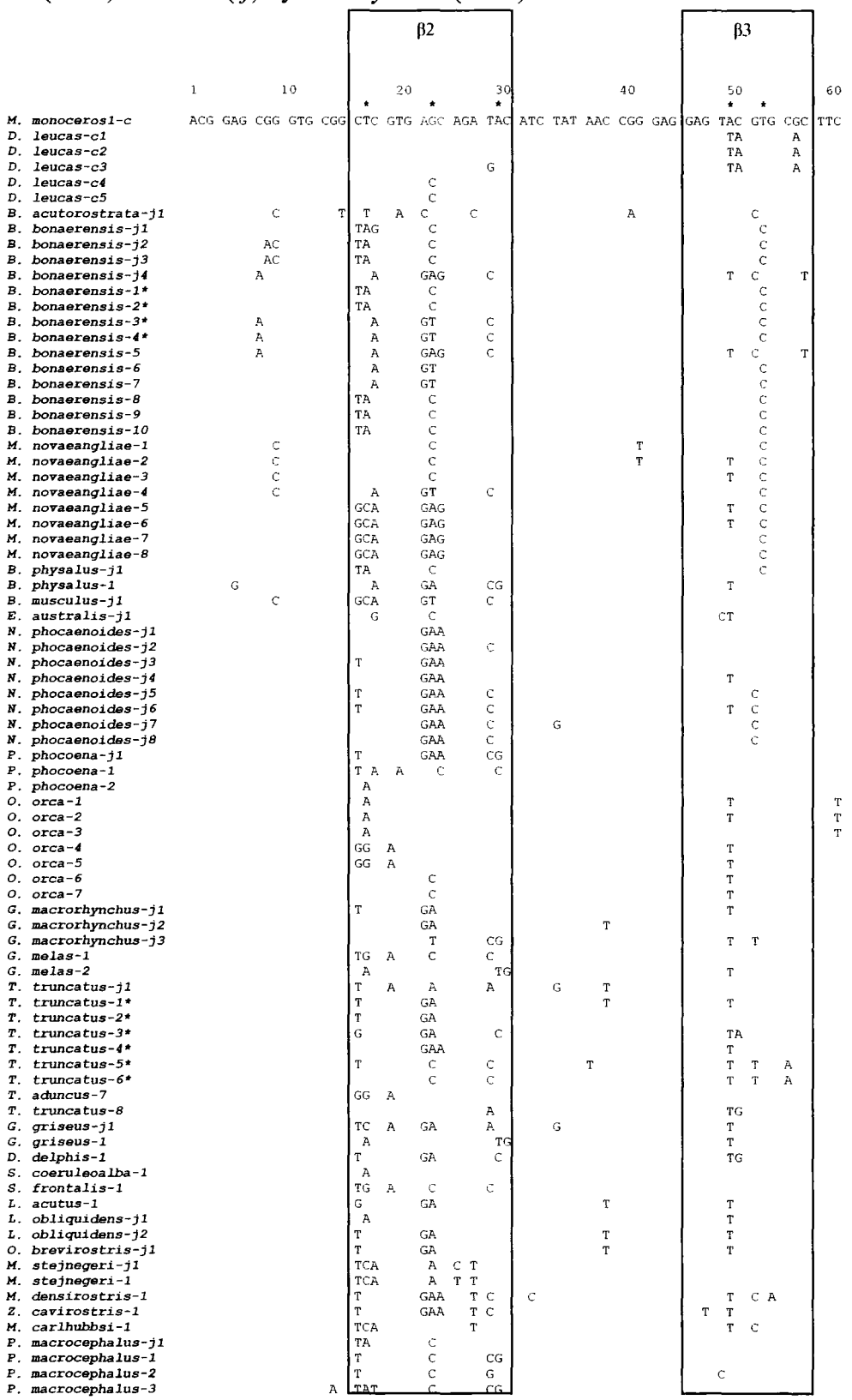


Figure 3.2.1: (continued)

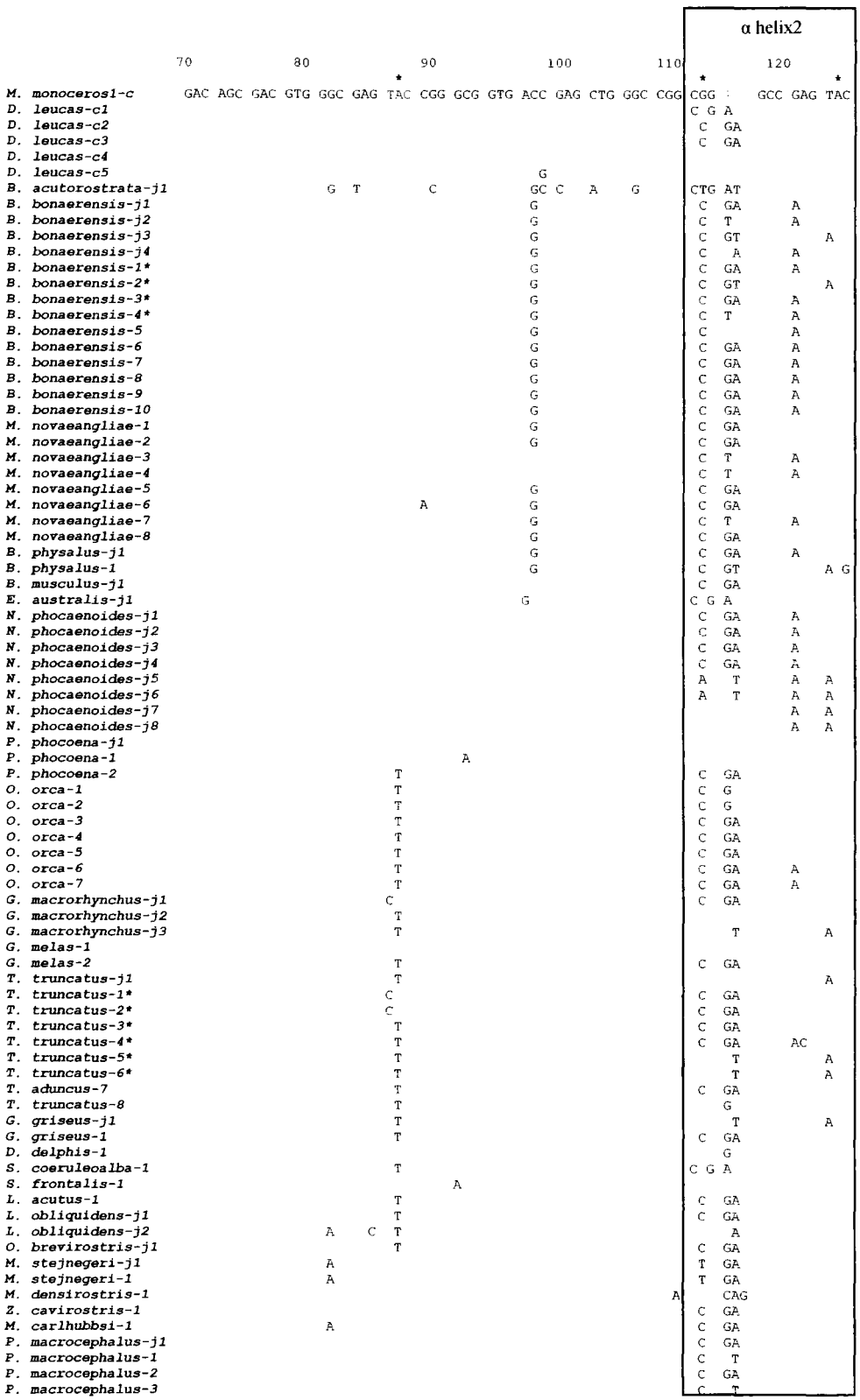


Figure 3.2.1: (continued)

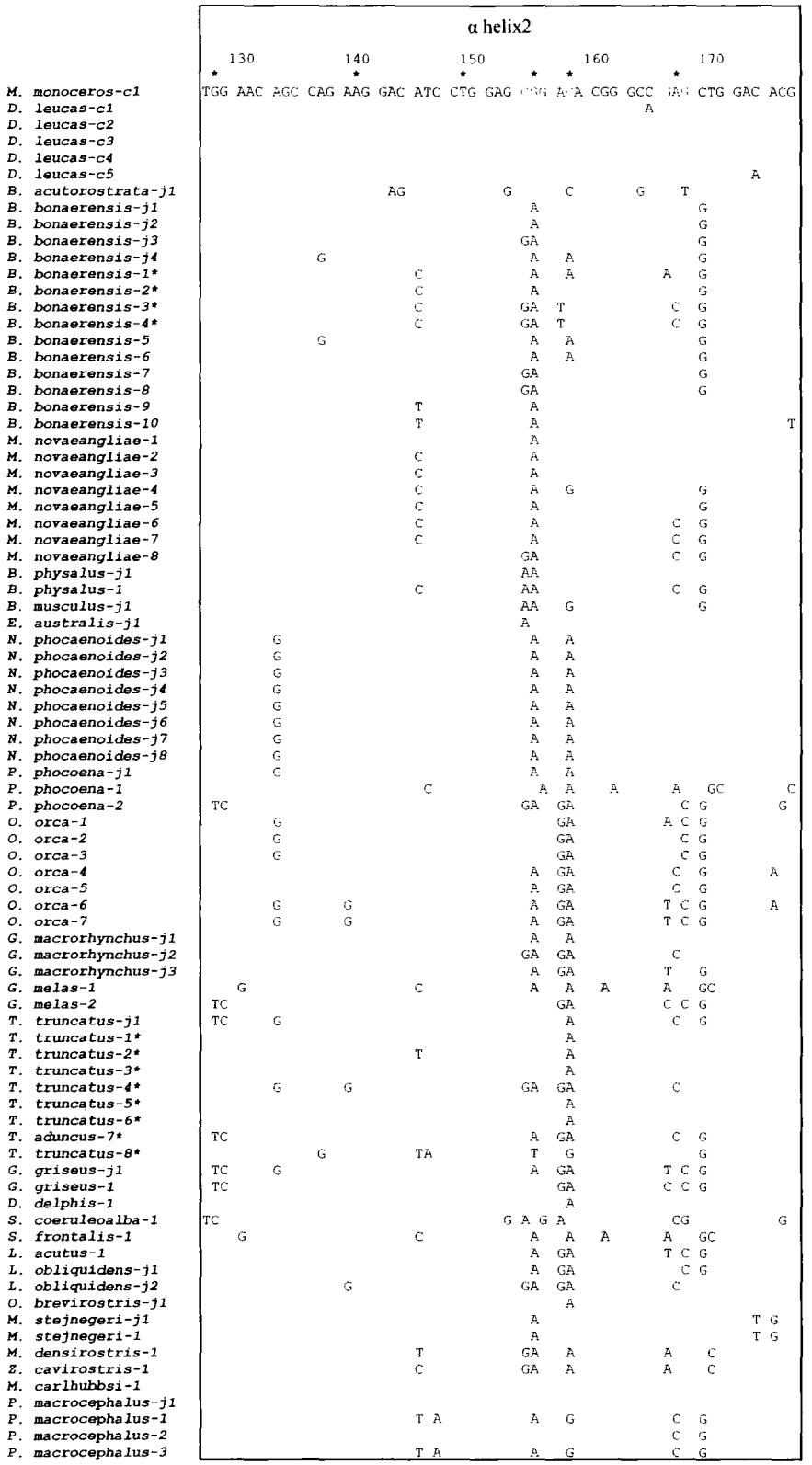


Figure 3.2.2: Amino acid sequence of the open reading frame of *DQB1* exon-2, based on the nucleotide sequence (Figure 3.2.1). Only the sites exhibiting nucleotide variation are shown. An asterisk denotes amino acids believed to be important in selective binding of different peptides and amino acids in bold font style polymorphic regions of HLA (Brown *et al.* 1993; Stern *et al.* 1994). Amino acid numbering is according to Brown *et al.* 1993. Sequences suffixed with (*-c*) were provided by Hayashi *et al.* (2003) and with (*-j*) by Murray *et al.* (1995).

```
                      21                                                         77
                           * * *       **         *         *   * *   *  *   **  *
M. monoceros-c1       TERVRLVSRY IYNREEYVRF DSDVGEYRAV TELGRRTAEY WNSQKDILER TRAELD
D. leucas-c1                           L H                  PD                  K
D. leucas-c2                           L H                  PD
D. leucas-c3                   D       L H                  PD
D. leucas-c4                 T
D. leucas-c5                 T                                                    E
B. acutorostrata-j1        Y T          A              S    PD K        F  Q
B. bonaerensis-j1            E H       FL              S    PN K     R     Q K   V
B. bonaerensis-j2       H  Y T          A              S    PS K           Q     V
B. bonaerensis-j3       H  Y T          A              S    PV  N          E     V
B. bonaerenses-j4          Y T          A              S    PD K           E     V
B. bonaerensis1*           Y T          A              S    PD K           E     V
B. bonaerensis2*           Y T          A              S    PD K        L  Q K  KV
B. bonaerensis3*             V H        A              S    PD K        L  E S  AV
B. bonaerensis4*             V H        A              S    PS K        L  E S  AV
B. bonaerensis5              E H       FL              S    P  K     R     Q K   V
B. bonaerensis6            Y T          A              S    PV  N       L  Q     V
B. bonaerensis7              V          A              S    PD K           Q K   V
B. bonaerensis8              V          A              S    PD K           E     V
B. bonaerensis9            Y T          A              S    PD K        F  Q
B. bonaerensis10           Y T          A              S    PD K           Q     V
M. novaeangliae1             V H        A                   PS K        L  Q R   V
M. novaeangliae2           A E         FA              S    PD          L  Q     V
M. novaeangliae3           A E         FA              S    PD          L  Q    AV
M. novaeangliae4             T          A              S    PD             Q
M. novaeangliae5             T         FA              S    PD          L  Q
M. novaeangliae6             T         FA                   PS K        L  Q
M. novaeangliae7           A E          A              S    PS K        L  Q    AV
M. novaeangliae8           A E          A              S    PD             E    AV
B. physalus-j1             Y T          A              S    PD K           K
B. physalus1           G     D R       F               S    PV  K       L  K    AV
B. musculus-j1             A V H                            PD             K R   V
E. australis-j1              T         L               S    PD             Q
N. phocaenoides-j1           E                              PD K    G      Q K
N. phocaenoides-j2           E H                            PD K    G      Q K
N. phocaenoides-j3         F E                              PD K    G      Q K
N. phocaenoides-j4           E         F                    PD K    G      Q K
N. phocaenoides-j5         F E H        L                   QI KN   G      Q K
N. phocaenoides-j6         F E H       FL                   QI KN   G      Q K
N. phocaenoides-j7           E H  D     L                      KN   G      Q K
N. phocaenoides-j8           E H        L                      KN   G      Q K
P. phocoena-j1             H                      F         PD    F        E E  DV
P. phocoena1               F E R                                G      Q K
P. phocoena2               YMT H                    T                   L  Q KQ KA
O. orca1                     T         F          F         PD K    G E    Q E  YV N
O. orca2                     T         F          F         PD K    G E    Q E  YV
O. orca3                   H           F          F         PA      G        E  NV
O. orca4                   H           F          F         PA      G        E  DV
O. orca5                   H                      F         PD      G        E  DV
O. orca6                   GM          F          F         PD             Q E  AV N
O. orca7                   GM          F          F         PD             Q E  AV
G. macrorhynchus-j1        F D         F          H         PD             Q K
G. macrorhynchus-j2          D                    F                        E E  A
G. macrorhynchus-j3          I R       FL         F          I  N          Q E  *V
G. melas1                  CMT H                                 D    L  Q KQ KA
G. melas2                  H   L       F          F         PD    F          E  HV
T. truncatus-j1            FMN N  D               F            N F G         K  AV
T. truncatus1              F D         F          H         PD               K
T. truncatus2              F D                    H         PD          F    K
T. truncatus3              V D S       L          F         PD               K
T. truncatus4              F T H   Y   FLS        F          I  N            K
T. truncatus5                T H       FLS        F          I  N            K
T. truncatus6              GM                     F         PD    F        Q E  AV
T. truncatus7                  N       L          F          A       R  Y  L R   V
T. truncatus8                E         F          F         PD T    G E    E E  A
G. griseus-j1              SMD N  D    F          F          I  N F G      Q E  YV
G. griseus1                H   L       F          F         PD    F          E  HV
D. delphis1                F D S       L                     A               K
S. coeruleoalba1           H                      F         PD    F        E E  DV
S. frontalis1              CMT H                    T              D    L  Q KQ KA
L. acutus1                 V D         F          F         PD             Q E  YV
L. obliquidens-j1          H           F          F         PD             Q E  DV
L. obliquidens-j2          F D         F        DDF          N        E    E E  A
O. brevirostris-j1         F D         F          F         PD               K
M. stejnegeri-j1           S N                  D           LD             Q
M. stejnegeri1             S NC                 D           LD             Q
M. densirostris1           F ESH T     FL                    Q          F  E K  KP
Z. cavirostris1            F ESH      DF                    PD          L  E K  KP
M. carlhubbsi-j1           S  S        FL        D          PD
P. macrocephalus1          F T D       H                    PD                  AV
P. macrocephalus-j1        Y T                              PD
P. macrocephalus2          F T R                            PI          L  Q R  AV
P. macrocephalus3         QY T R                            PI          L  Q R  AV
```

Figure 3.2.3: Maximum Likelihood estimation of site rates in *DQB1* exon-2 sequence.

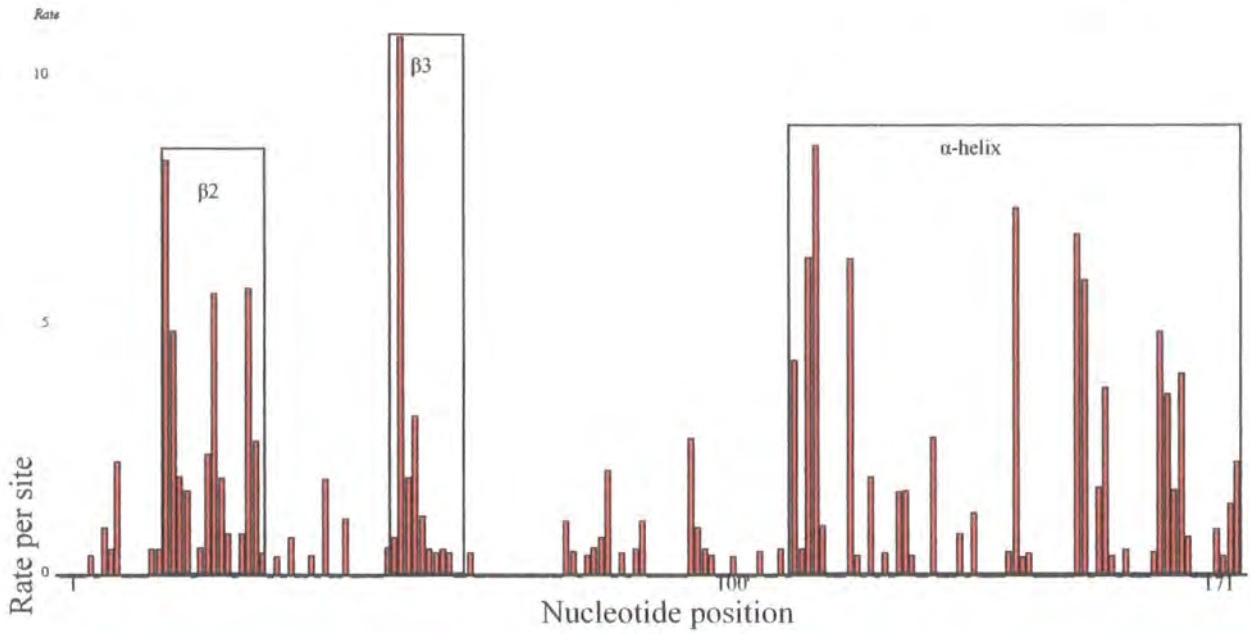
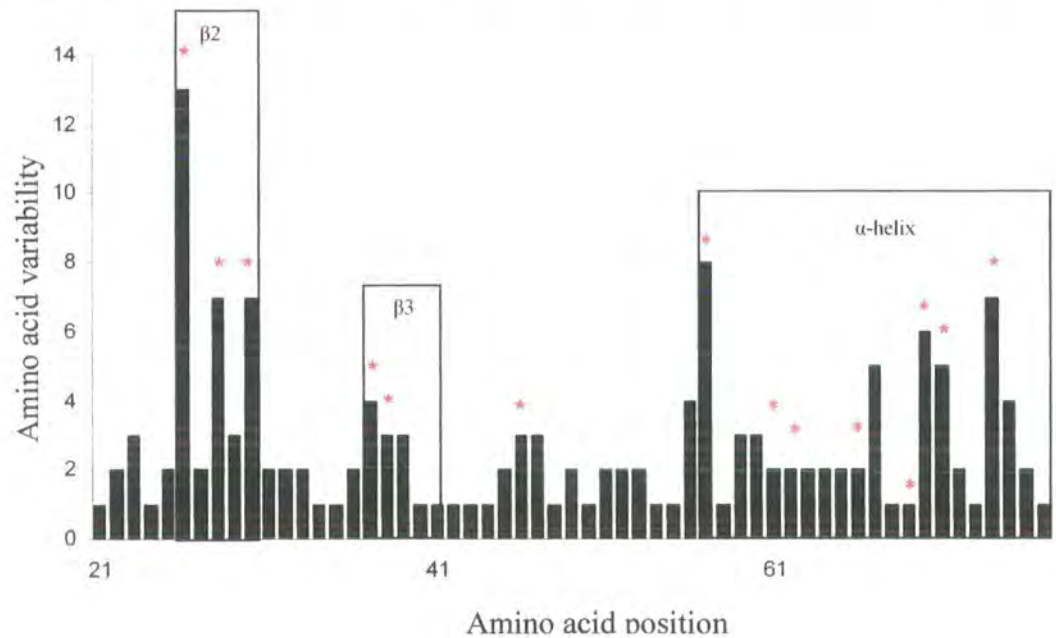


Figure 3.2.4: Variability plot for *DQB1* exon-2. The boxes indicate protein topology for beta chains and alpha helix. Red asterisks above bars indicate putative PBR (Brown *et al.* 1993).



3.2.1.2 Amino acid substitution patterns

I investigated two different parameters with respect to amino acid substitutions. The first is an estimation of the relative number of synonymous (silent) and non-synonymous (amino acid altering) nucleotide substitutions, while the second assesses the physicochemical properties of the substitution, such as charge, polarity and volume. It has been indicated that the nonsynonymous/ synonymous substitution rate ratio positively correlates to the radical/ conservative rate ratio (Zhang 2000; Smith 2003).

Nonsynonymous/ synonymous substitutions

Table 3.2.2 summarizes the results of the four different methods used to obtain the nucleotide substitution patterns in exon-2 (see Chapter 2). Both the approximate and the maximum likelihood methods show that the *dn/ds* ratio is significantly greater than one in the PBR sites, an observation consistent with studies in other mammal species and an indication of selection (range of t-values=3.647-9.892, p-values=0.002-<0.001, df=26). The highest ω values are observed in the PBR sites whilst the lowest in the non-PBR sites and this indicates that selection is stronger in the putative PBR sites. In addition, the ω ratio was not significantly different than one in the non-PBR sites.

The SLAC phylogeny (see Chapter 2: 2.6.2.4) in Figure 3.2.5 indicates that distances between non-synonymous substitutions are smaller within than between cetacean families and this result in the clustering of species of the same family. In contrast, according to the synonymous substitutions in Figure 3.2.6 phylogeny species are all clustered in the center, indicating even phylogenetic distances between different species and families. The results of the SLAC phylogenetic analysis hypothetically suggest that ecological differences exhibited by different cetacean families may also affect the selection pressure and evolution of *DQB1* locus.

The closer the gamma parameter value (α) is to zero, the higher the selection, as α is inversely proportional to the coefficient of variation. The GAMMA program (Gu and Zhang 1997) was used in order to estimate the gamma parameter of the exon-2 sequences identified in this study (see Chapter 2: 2.6.1.5). The value was α=0.34, an observation which supports the hypothesis that exon-2 is under positive selection (Smith 2003). In addition, the MEGA version 2.1 (Kumar *et al.* 2001) program was used to estimate the transition over transversion ratio (κ), and κ=2.53 (see Chapter 2). The κ value supports the hypothesis that exon-2 refutes neutrality as DNA sequences under neutrality exhibit κ ratios close or equal to one (Kumar *et al.* 2001).

Table 3.2.2: Number of synonymous substitutions per synonymous site (*ds*) and number of nonsynonymous substitutions per nonsynonymous site (*dn*) in the full length of the exon-2 sequence and the putative non-PBR and PBR sites respectively. Numbers in brackets indicate Standard Error. The *dn/ds* ratio is denoted by ω. An asterisk denotes values significantly different from one.

	Full length	PBR	non-PBR
Nei-Gojobori			
S	48.515	11.837	36.678
N	111.506	27.163	84.343
ds	0.0292 (0.0076)	0.0233 (0.0144)	0.0302 (0.0094)
dn	0.1041 (0.0234)	0.3008 (0.0823)	0.0536 (0.0173)
ω	3.5651*	12.9099*	1.7748
Li *et al* (1985)			
S	56.462	4.280	21.624
N	104.670	27.042	77.602
ds	0.0821 (0.0196)	0.0603 (0.0270)	0.0475 (0.0150)
dn	0.1039 (0.0232)	0.2781 (0.0778)	0.0519 (0.0170)
ω	1.2655	4.6119*	1.0926
Yang-Nielsen (ML)			
S	58.309	16.714	28.452
N	112.692	25.286	90.547
ds	0.106	0.076	0.102
dn	0.251	0.422	0.141
ω	2.365	5.549*	1.380
Kosakovsky *et al.* (2004)			
S	50.374	17.655	31.708
N	120.626	23.449	93.842
ds	0.082	0.085	0.085
dn	0.263	0.458	0.097
ω	3.197*	5.431*	1.134

The simplistic assumptions which are made by the Nei-Gojobori (1986) method tend to overestimate the non-synonymous and synonymous sites in the sequences because this method ignores transition/transversion ratio and codon usage. As a consequence ω has high values compared to the other methods; in the PBR in particular. The Li *et al.* (1985) method takes into consideration the κ and α values and thus, ω values obtained are smaller than the values obtained from the Nei-Gojobory method. However, the codon-usage bias in Li's method results in the overestimation of synonymous changes in synonymous sites and due to this assumption, the ω value in full length and in non-PBR sites is smaller than the values obtained form all the other methods(seeTable3.2.2).

Figure 3.2.5: SLAC phylogeny of the synonymous substitutions of all the *DQB1* alleles included in this study

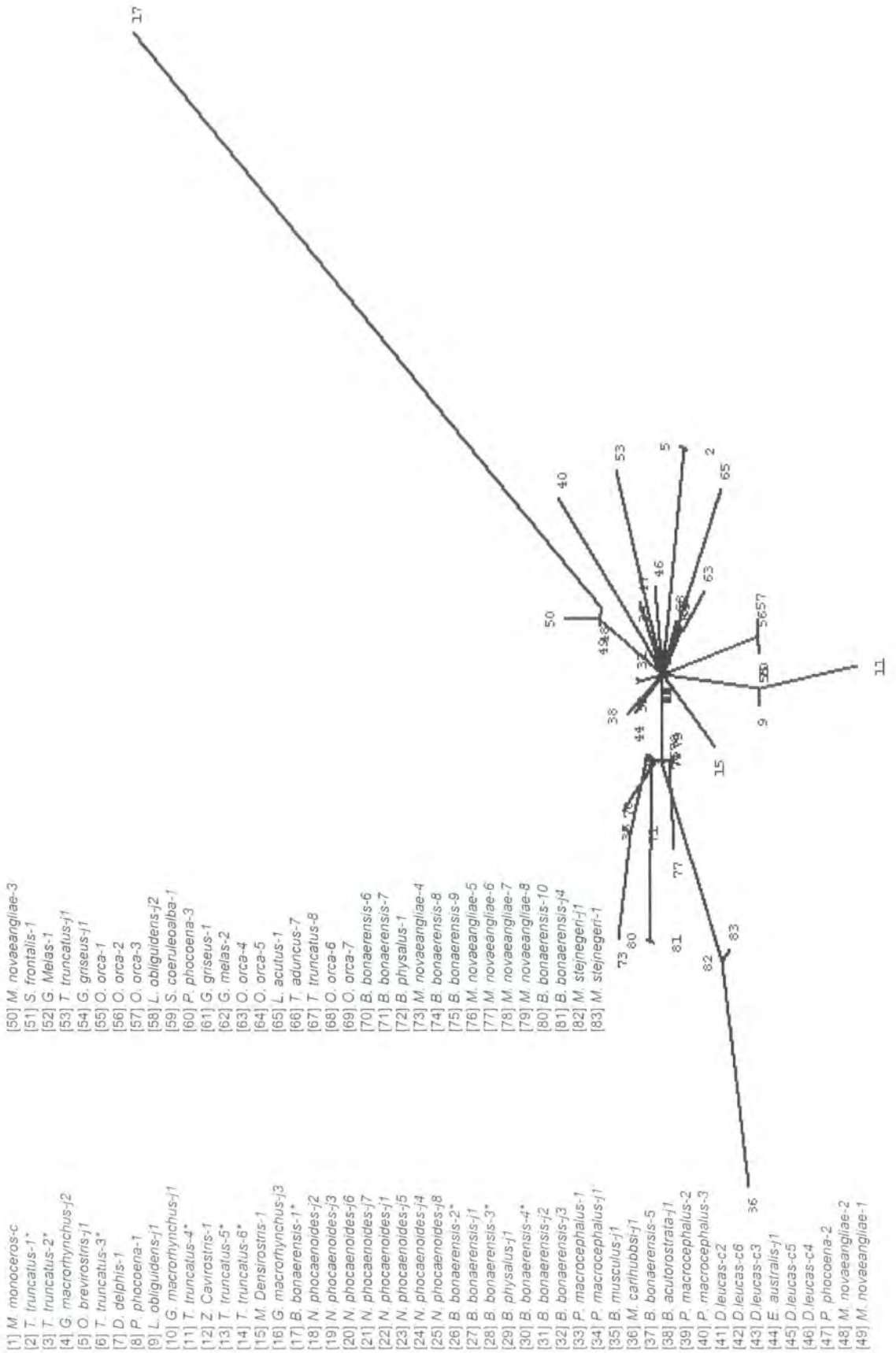
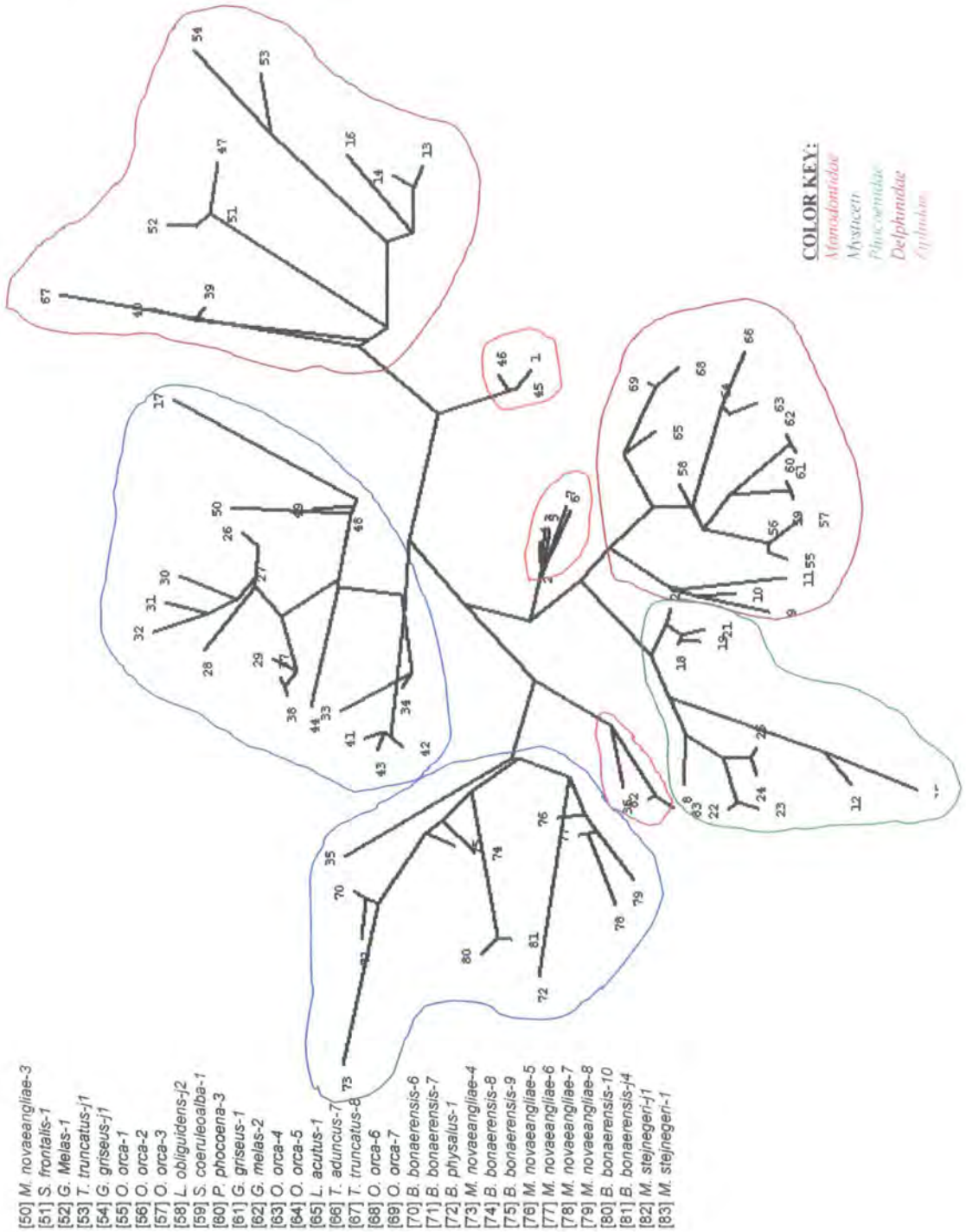


Figure 3.2.6: SLAC phylogeny of the non-synonymous substitutions of all the *DQB1* alleles included in this study.



The ML methods are based on explicit models of codon substitution and all the different parameters in the model *i.e.* sequence divergence, transition/transversion ratio and the *dn/ds* ratio are estimated from the data. The analysis for the ML methods was performed at three different ω initial values: ω=1, ω=2 and ω=0.5 and the analysis which exhibited the highest likelihood value was used. In addition the model for codon substitution was derived from the present data in order to accurately estimate the ω value.

Radical/ conservative substitutions

Figure 3.2.7 shows the ratio of radical over conservative amino acid substitutions in exon-2 for each allele identified in the present study (see Chapter 2). A significantly higher rate of radical substitutions has been taken as evidence for positive selection, even when there is no observation of significantly higher rate of non-synonymous than synonymous substitution (Gu and Zhang 1997; Zhang 2000; Smith 2003).

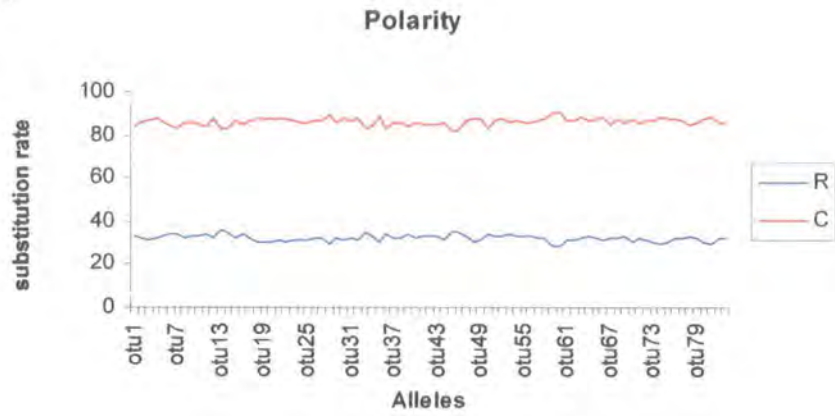
The radical/ conservative ratios in polarity (Figure 3.2.7-A) and charge (Figure 3.2.7-B) is 0.284 and 0.952 respectively, which is not significantly different than one (t-value=; p-value=; df=). This indicates that exon-2 is under neutrality, as it is only the radical substitutions which alter the physicochemical properties and thus having an effect on protein functions (Gu and Zhang 1997). However, Zhang (2000) showed that polarity and charge alone should not be considered as accurate properties in predicting functionally important substitutions.

Previous studies have shown that the intensity of positive selection is not significantly greater for radical changes than for conservative changes when only charge or polarity is considered (Gu and Zhang 1997; Zhang 2000). Therefore, Zhang (2000) suggested that the combined amino acid analysis of polarity and volume will give an accurate estimation of positive selection. These observations are consistent to the results of the present study.

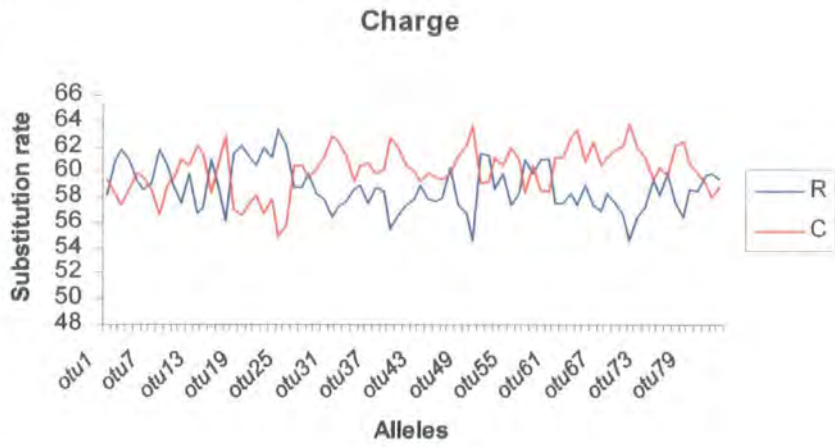
In Figure 3.2.7 A and B, where charge and polarity are considered individually, the ratio of radical/ conservative substitutions indicates neutrality; however, in Figure 3.2.7-C where the properties of both polarity and volume are considered the ratio is 3.074, significantly higher than one (t-value=3.471; p-value=0.003; df=164). This may further suggest that *DQB1* exon-2 in cetaceans is under positive selection.

Figure 3.2.7: Numbers of conservative and radical substitutions per site for cetacean *DQB1* exon2 considering: A) Polarity, B) Charge and C) Polarity and volume (Miyata *et al.*1979).

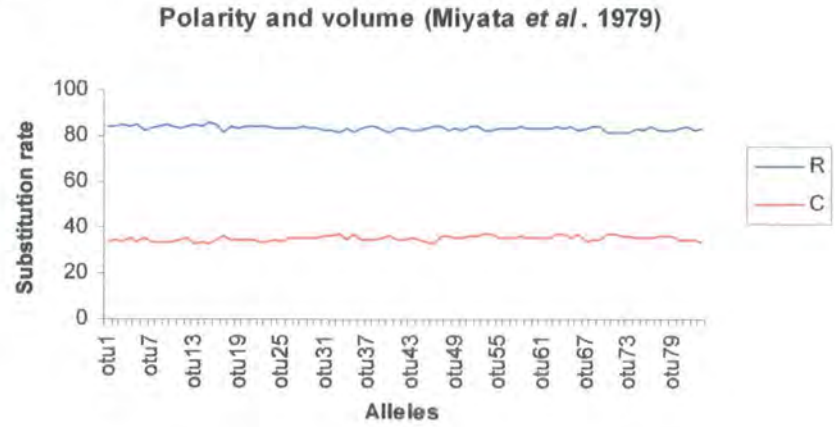
A)



B)



C)



3.2.2 *DQB1* Phylogenetics

The phylogeny of cetacean *DQB1* alleles was investigated under three different phylogenetic models: a) Tamura-Nei (TN; Figure 3.2.8), b) Maximum Likelihood (ML; Figure 3.2.9) and c) Maximum Parsimony (MP; Figure 3.2.10); all trees were based upon 1000 bootstrap replications. The Tamura and Nei model (1993) takes into account transition/ transversion bias, gamma parameter of substitution rates (α) and nucleotide frequencies. Similar results were obtained by all different methods indicating that the patterns of sequence relationships are independent of the method used for phylogenetic reconstruction. To test that convergent evolution was not responsible for the observed branching order, the TN analysis was performed by taking into consideration the PBR and the non-PBR sites individually (Figures 3.2.12 and 3.2.13 respectively). It is indicated that the similarity between alleles is due to common ancestry, as both trees exhibit similar pattern of phylogenetic relationships. The phylogenetic analysis of the cetacean *DQB1* alleles identified in this study reveals a trans-species mode of evolution where alleles are clustered according to types rather than to species. Figure 3.2.11 shows a TN tree of cytochrome-*b* in order to illustrate the trans-species evolution pattern exhibited by MHC evolution.

Figure 3.2.8 illustrates the Neighbor-Joining tree of exon-2 of *DQB1* locus constructed by the Tamura-Nei distance estimation model, which takes into consideration transition/ transversion and nucleotide frequencies estimated from the data using the Nei-Gojobori method with Jukes-Cantor correction (see Chapter 2: 2.6.1.1). Bootstrap values range between 51–100%. Thus, the phylogeny is strongly supported. Two HLA *DQB1* sequences are used as outgroups. *DQB1* sequences of ungulates, which show a homology of up to 92%, are not included in the cetacean lineage. The cetacean phylogeny is characterized by an extended polytomy with short branch lengths. Trans-species lineages are formed, for example: 1) *P. phocoena-1, S.coeruleoalba-1, G. melas-1, G. griseus-1* (bootstrap value 55), 2) *P. phoconea-1, G. melas-2, S.frontalis-1, L. vexillifer-4,-5* (bootstrap value 88), 3) *M. densirostris-1, Z. cavirostris-1* (bootstrap value 95).

Figures 3.2.9 and 3.2.10 show a ML and MP phylogeny respectively. The range of the bootstrap values is between 50–99% in both trees. Although the cetacean sequences exhibit similar branching pattern as in the TN phylogeny, the main difference is that the *DQB1* sequences from the ungulates falls within the cetacean lineage.

Figure 3.2.8: Neighbor-Joining tree of *DQB1* exon2 sequences identified in this study. Tree is based upon 1000 bootstrap replications. Numbers in bold font style above branches indicate bootstrap value, numbers below in regular style the distances.

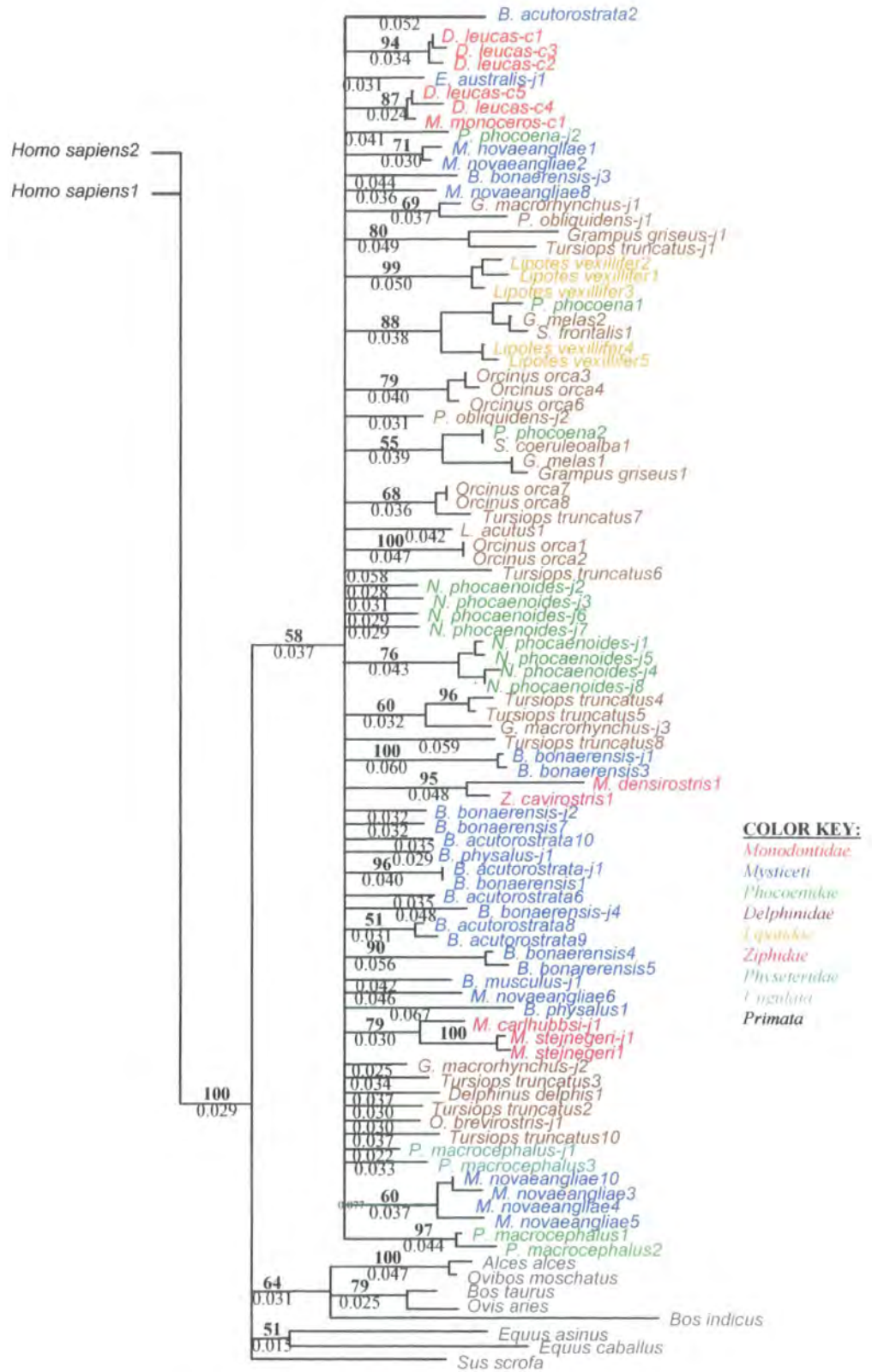


Figure 3.2.9: Maximum Likelihood tree of *DQβ* exon2 sequences identified in this study. Tree is based upon 1000 bootstrap replications. Numbers in bold font style above branches indicate bootstrap value, numbers below in regular style the distances.

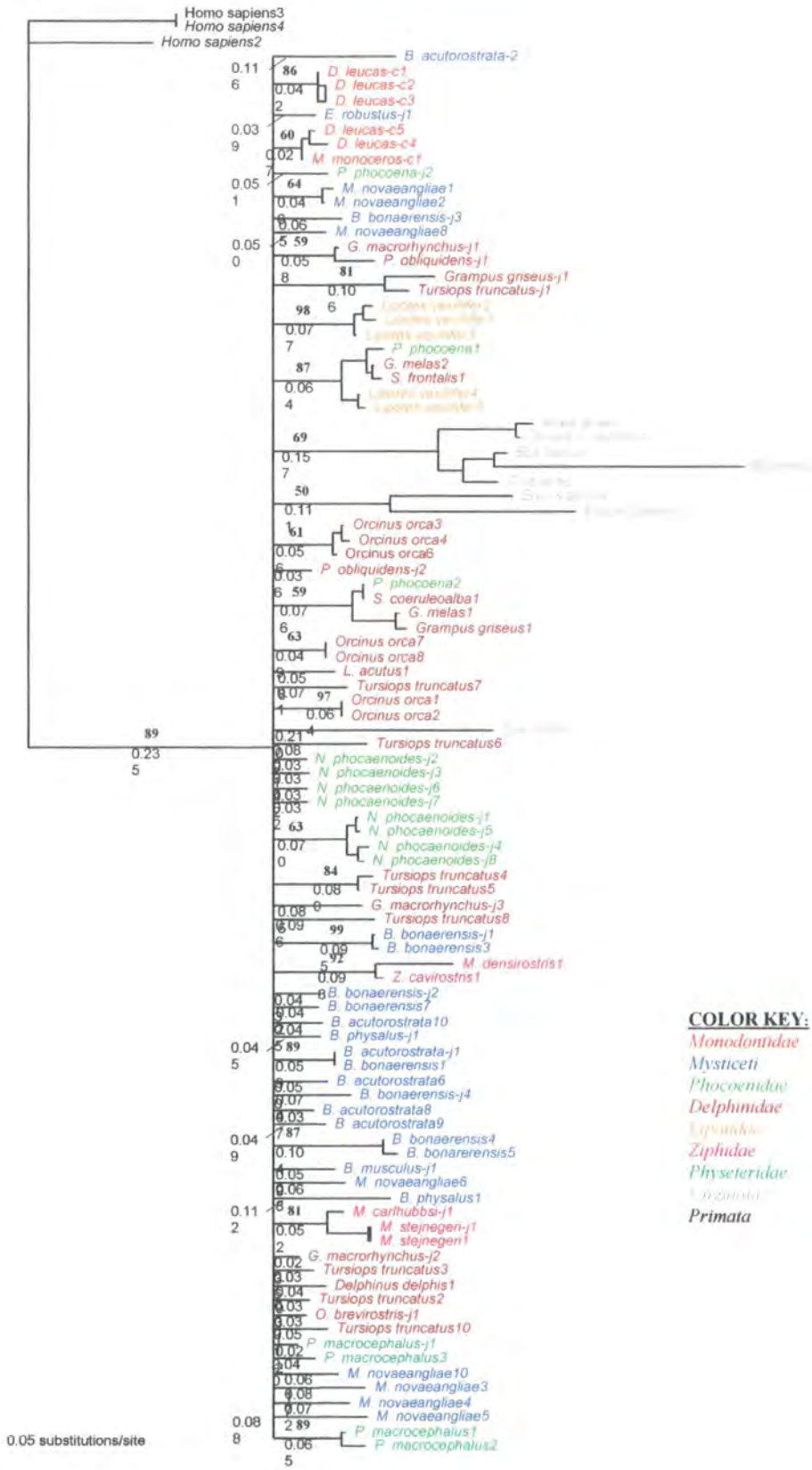


Figure 3.2.10: Maximum parsimony tree of *DQβ* exon2 sequences identified in this study. Tree is based upon 1000 bootstrap replications. Numbers in bold font style above branches indicate bootstrap value, numbers below in regular style the distances.

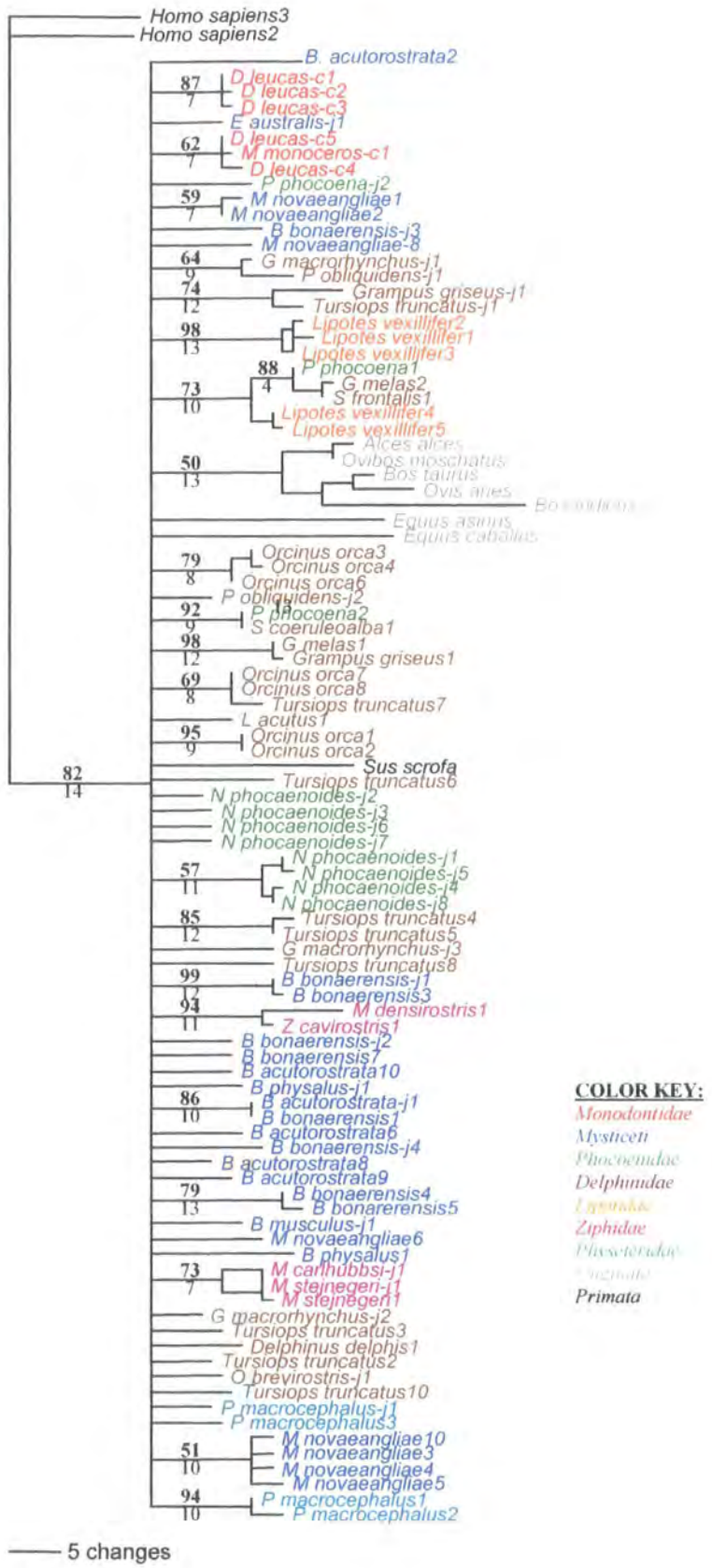


Figure 3.2.11: Neighbor-Joining tree of cytochrome - *b* sequences of all the cetacean species used in this study. Tree is based upon 1000 bootstrap replications. Numbers in bold font style above branches indicate bootstrap value, numbers below in regular style the distances.

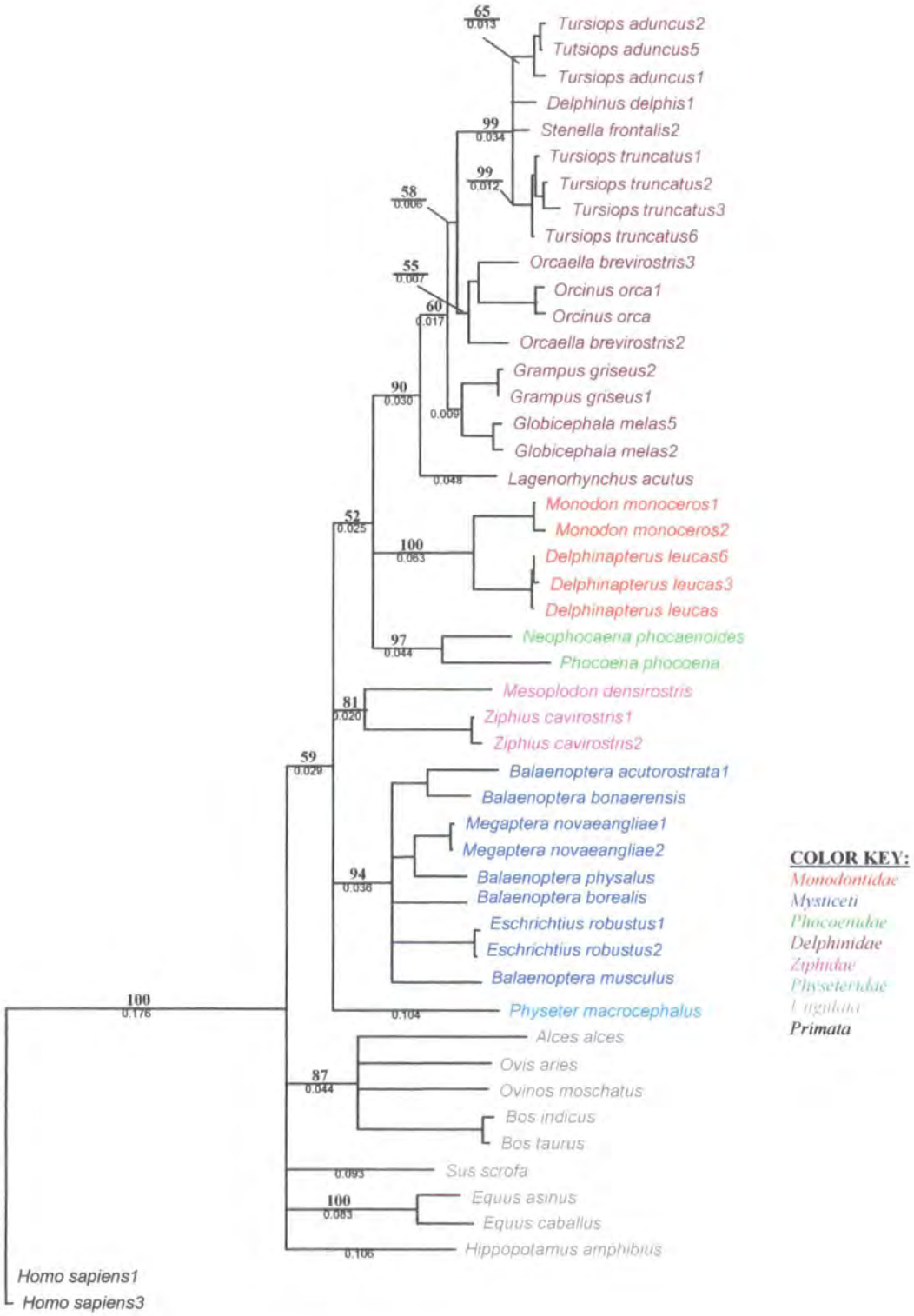


Figure 3.2.12: Neighbour Joining tree of *DQB1* exon-2 putative PBR sites. Tree is based upon 1000 bootstrap replications. Numbers in bold font style above branches indicate bootstrap value, numbers below in regular style the distances.

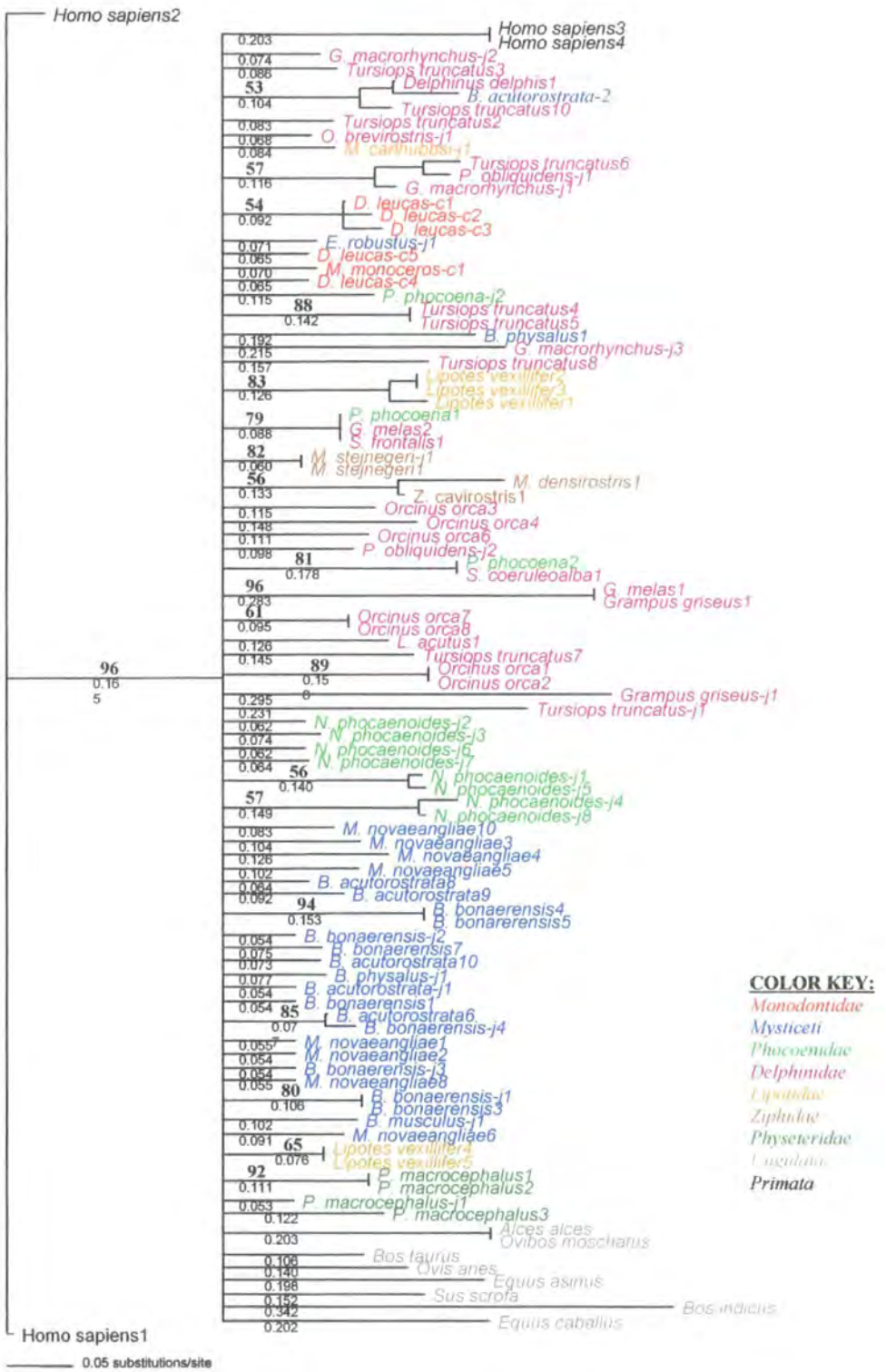
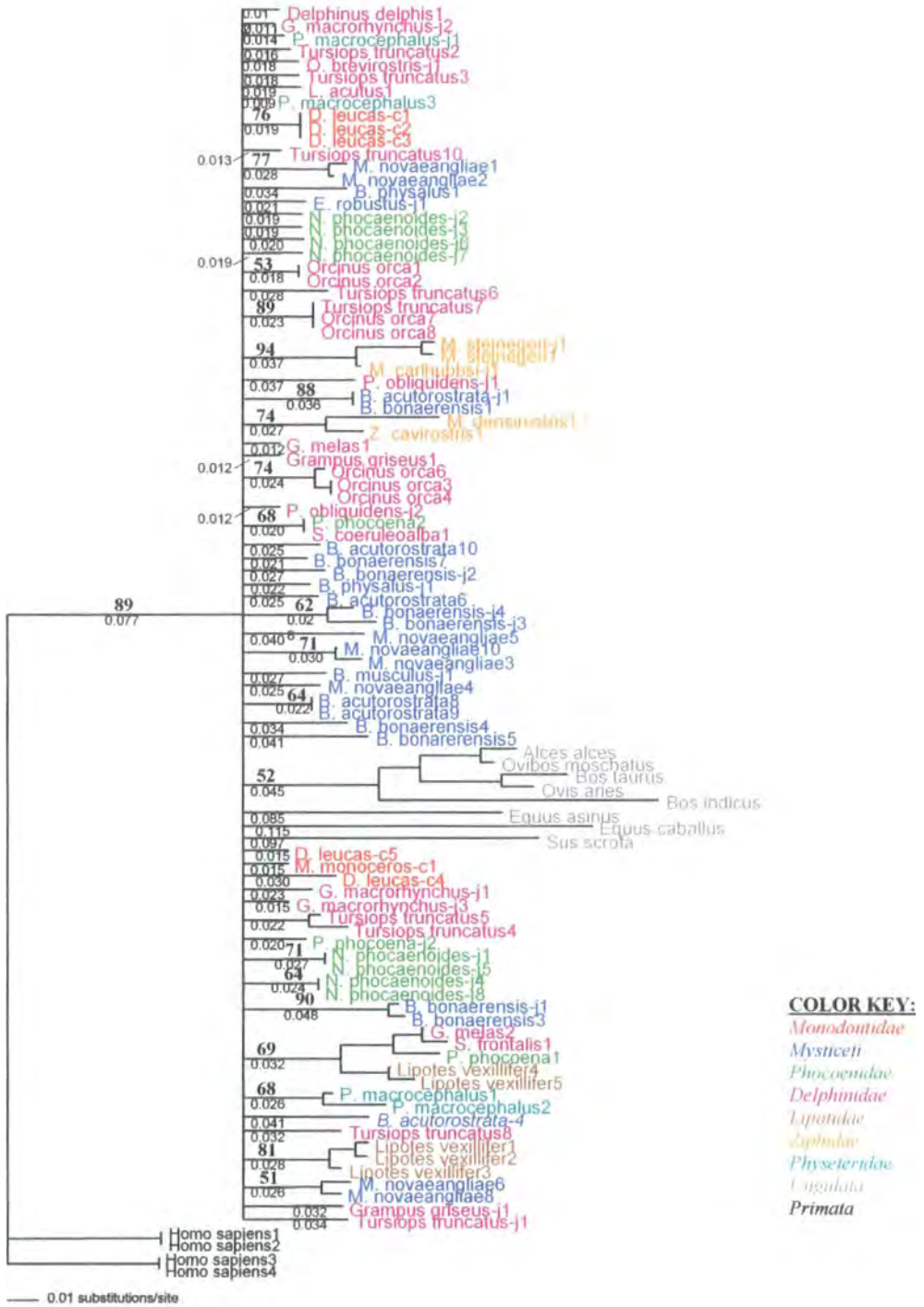


Figure 3.2.13: Neighbour Joining tree *of DQB1* exon-2 putative non-PBR sites. Tree is based upon 1000 bootstrap replications. Numbers in bold font style above branches indicate bootstrap value, numbers below in regular style the distances.



The cetacean sequences again show the same topology as in the previous phylogenies. Ungulates are included in the cetacean lineage.

Figure 3.2.12 shows the phylogeny of the putative PBR sites in cetacean *DQB1* alleles while figure 3.2.13 shows the phylogeny of the non-PBR sites. In the PBR phylogeny, there are seven allelic clusters denoting trans-species evolution and the branch length of the cetacean sequences denote at least 0.05 substitutions per site. In the non-PBR phylogeny there are two clusters, both within *Ziphidae*, which indicate trans-species evolution and branch lengths denote 0.01 substitutions per site.

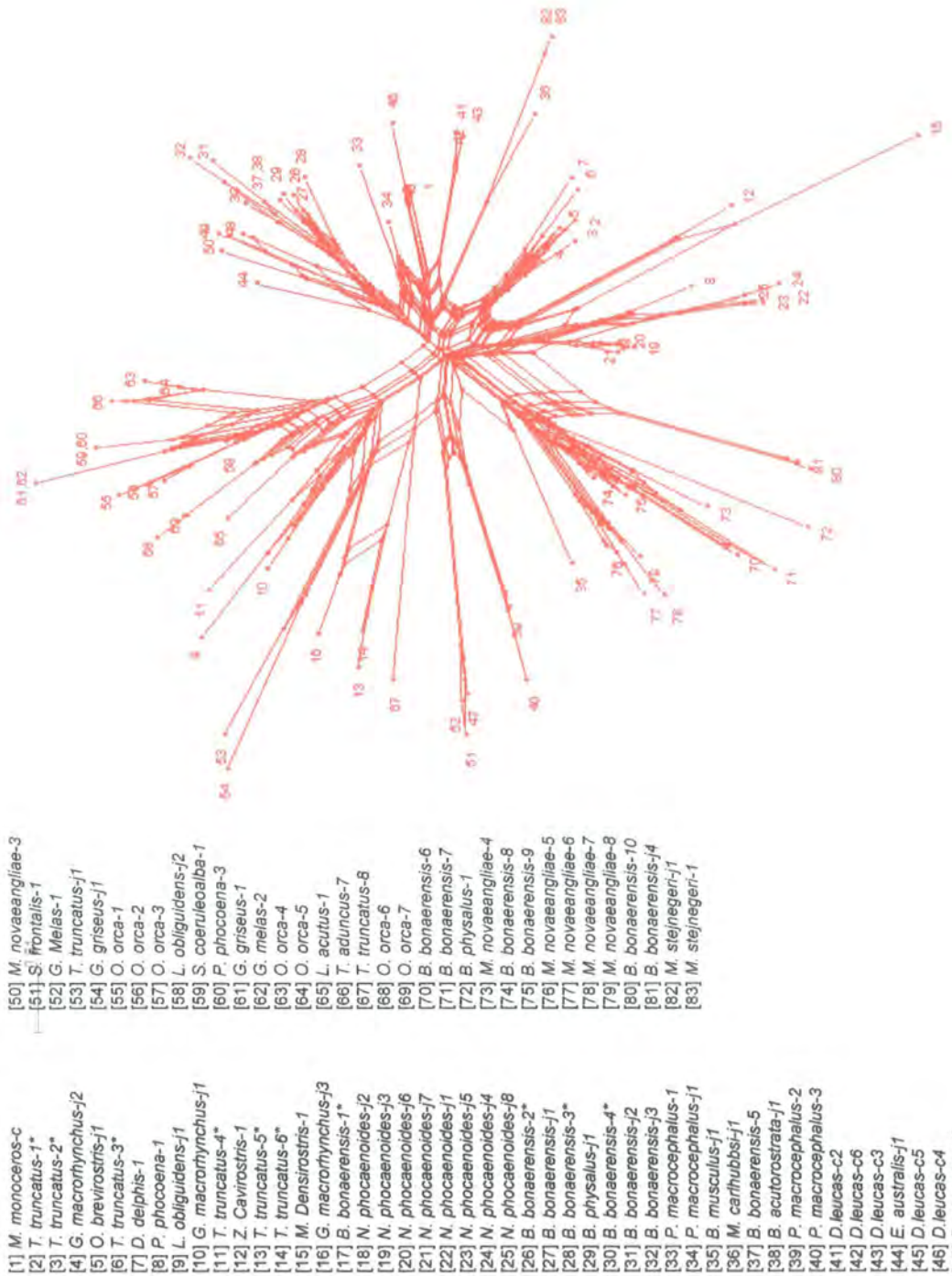
3.2.2.1 Recombination

Fitch (1997) showed that recombination produces networks of sequences rather than strictly bifurcating trees, therefore, the initial search for recombination in MHC *DQB1* exon2 was conducted using split decomposition (see Chapter 2; Huson 1998). This analysis depicted parallel edges between sequences if there are conflicting phylogenetic pathways in the data and it was undertaken with the SplitsTree program (version 2.4; Hunson 1998). Figure 3.1.14 shows that *DQB1* alleles are linked to each other by multiple pathways, suggesting evidence for networked evolution. In order to determine which of the alleles were the most likely recombinants, each nucleotide site was mapped onto the splits graph (see Chapter 2: 2.6.3; Holmes *et al.* 1999). This analysis revealed that the alleles *G. melas-2* and *B. acutorostrata-2* contained distinct regions which showed closest connections to different alleles in the sample. *G. melas-2* sequence closely matched that of *P. phocoena-1* and *S. frontalis-1*and *B. acutorostrata-2* closely matched that of *B. physalus-1* and *B. bonaerensis-j4*. LARD analysis (version 1; Holmes *et al.*, 1999) was performed, in order to map individual recombination event (referred to as break-point) on the alleles *G. melas-2* and *B. acutorostrata-2* (see Chapter 2). This analysis indicated that the putative break-point in *B. acutorostrata-2* is at site 30 (Figure 3.2.1) and in *G. melas-2* at site 87.

3.3 Discussion

This study evaluated the radiation and the level of polymorphism in the *DQB1* locus in the mammalian order *Cetacea*. The pattern of evolution and the relatively

Figure 3.2.14: Split decomposition analysis of all the *DQB1* alleles included in this study.



high levels of polymorphism for the cetacean *DQB1* are consistent with observations for several other mammalian orders, for which the pattern has been attributed to natural selection (Gyllesten *et al.* 1990; Hughes and Yeager 1998; Seddon and Ellegren 2002). Previous studies in cetaceans (see 3.1) have suggested relatively low levels of MHC variation. However, present results argue that cetaceans exhibit similar levels of variation as terrestrial mammals. I suggest that MHC *DQB1* variation in cetaceans is comparable to terrestrial mammals for the following reasons.

Two of the main characteristics of MHC polymorhism attributed to natural selection are the trans-species evolution pattern and high non-synonymous substitutions rates (Klein 1986; Hudges and Nei 1988,1989; Wills 1991; Hughes and Yeager 1998; Klein *et al.* 1998). Results of the evaluation of *DQB1* evolution in cetaceans (Figures 3.2.8-3.2.10) demonstrated that those two characteristics are observed in the cetacean *DQB1* radiation.

In a study of allelic diversification at the *DQB1* locus in the order *primata*, Gyllesten *et al.* (1990) compared 30 humans of different ethnical origin (Caucasian, African and Asian), 13 chinpanzees (*Pan troglodytes*), 4 pygmy chimpanzees (*Pan paniscus*), 6 gorilas (*Gorilla gorilla*), 1 baboon (*Papio leucapheus*), 2 rhesus monkeys (*Macaca mullata*), 2 langurs (*Presβytis entellus*) and 1 capuchin monkey (*Ceβus capuchinus*) and found 38% of the *DQB1* locus lineages to exhibit trans-species evolution. The same study found a mean of nonsynonymous to synonymous (*dn/ds*) ratio of 2.05 (Nei-Gojobori method, full length of exon-2). Hughes and Yeager (1998) showed that 53% of *DQB1* allelic lineages between human and chimpanzee (*Pan troglodytes*) exhibit trans-species polymorphism and a *dn/ds* ratio of 2.643 (Nei-Gojobori method, full length of exon-2). In a study on *DQB1* exon-2 evolution in the family *canidae*, Seddon and Ellegren (2002) compared European (n=163) and North American (n= 8) wolves (*Canis lupus*), dogs (n=30; *Canis fascicularis*) and coyote (*Canis latrans*) and found 42% of allelic lineages to exhibit trans-species polymorphism. The *dn/ds* ratio was estimated as 2.43 (Nei-Gojobori method, full length of exon-2). The present study found that in the cetacean order, 38% of *DQB1* allelic lineages exhibit trans-species polymorphism and the *dn/ds* ratio is 3.565 (Nei-Gojobori method, full length of exon-2; Table 3.2.2). Furthermore, the phylogenetic relationship among alleles of the *DQB1* locus in primates and canines showed similar patterns to that seen in the *Cetacea* order. In each of the cases the phylogenetic trees inferred exhibit extended polytomies and low levels of taxonomic structure.

For primate and canine studies, trans-species polymorphism and nucleotide substitution patterns have been attributed to natural selection by the pathogenic pressure of the terrestrial environment. Cetaceans exhibit comparable patterns of trans-species polymorphism and nucleotide substitution in the marine environment suggesting a similar level of selection in response to marine pathogens.

3.3.1 Conclusions

The observed pattern of evolution of MHC class II *DQB1* allele in the cetacean species involved in this study indicates that exon-2 is under selection pressure and variation is being maintained by positive selection.

Two main conclusions are drawn through the phylogenetic analysis: a) phylogenetic relationship among the cetacean species does not recapitulate the expected taxonomic relationship, b) alleles from different species and different genera fall in the same allelic lineages. In the cytochrome-b NJ tree there is a clear distinction between *Odontoceti* and *Mysticeti* and at the family level as well. The cytochrome-*b* NJ tree shows a characteristic evolutionary pattern of a functional protein and upon comparison with *DQB1*, the different evolutionary pattern is obvious. An extended polytomy and a trans-species evolutionary pattern are characteristics for the MHC genes in mammalian species and it is the same pattern observed in this study.

Sequence analyses show that cetacean *DQB1*: a) is implicated in the immune response of cetaceans and is functional and b) is selected positively and under selection pressure. Higher number of non-synonymous to synonymous substitutions is expected for MHC loci as they are under the influence of balancing selection (see chapter 6). In addition, the putative PBR sites identified from HLA analysis exhibited the highest non-synonymous to synonymous substitution rates compared to non-PBR and to full length analyses.

CHAPTER 4:

EFFECTS OF CETACEAN LIFE HISTORY ON *DQB1* EVOLUTION

4.1 INTRODUCTION AND OBJECTIVES

Studies in disease dynamics and epidemiology in mammals indicate that social behaviour can affect the disease transmission rate (Begon *et al.* 2002; Day and Burns 2003; Mylius and Metz 2004) and that social species are under higher pathogen pressure than less social species (Bowers and Turner 1997; Beardmore *et al.* 2001; Read and Keeling 2003; Gudelj and White 2004). In the *Odontoceti* suborder, at least 89% of the extant species are social and form groups, living at very close proximity to each other in groups, while, at least 93% of *Mysticeti* species are solitary (Connor 2002; Martin and Reeves 2002). The divergence between *Odontoceti* and *Mysticeti* in the cetacean order took place approximately 33-34 million years ago (Thewissen 1998; Nikaido *et al.* 1999), early in the radiation of *Cetacea*, and thus allowing for possible differences in selection pressure to affect the pattern of MHC evolution in each suborder respectively.

4.1.1 Cetacean social organization

4.1.1.1 *Odontoceti* social behaviour

Two main social structures have been observed in *Odontoceti* species, a consistent, stable society, and a flexible fluid society (Connor *et al.* 1998). The former has been observed in the pilot whales (*G. macrorhynchus* and *G. melas*; Olson and Reilly 2002), killer whale (*O. orca*; Ford 2002), sperm whale (*P. macrocephalus*; Whitehead and Weilgart 2000); the latter in the bottlenose dolphins (*T. aduncus* and *T. truncatus*; Connor *et al.* 1999; Scott and Wells 2002), Risso's dolphin (*G. griseus*; Baird 2002), common dolphin (*D. delphis*; Evans 1994), Atlantic spotted dolphin (*S. frontalis*; Carwardine 1995), Striped dolphin (*S. coeruleoalba*; Archer and Perrin 1999),

Atlantic white sided dolphin (*L. acutus*; Leduc 2002), harbor porpoise (*P. phocoena*) and finless porpoise (*N. phocaenoides*) among others (Read 2002).

In the killer whale, social associations are formed in small matrilineal groups (referred to as pods) with an average size between 15-20 individuals (Bigg *et al.* 1990). Pilot whale society is believed to be very similar to that of the killer whale, where large groups of up to 200 individuals are comprised of smaller pods defined by strong bonds among individuals (Olson and Reilly 2002). In sperm whales, females form stable associations of approximately 12-15 individuals (Whitehead 2000). Males upon maturity leave the unit to form loose aggregations (referred to as 'bachelor groups') with males of similar age and size (Whitehead 2000). As male sperm whales age, social associations and group size tend to diminish (Whitehead 2002).

The fluid society system is exhibited by the smaller members of the suborder, where associations are generally brief and groups composition dynamic (Connor 2002; Leduc 2002). A characteristic example is the social structure of the common dolphin where a large group of up to hundreds of individuals is formed composed of smaller more stable groups (Evans 1994). Nevertheless, the structure of each of the smaller groups is unstable and temporary, compared to 'stable society' species, such as the killer whale (Leduc 2002).

Beaked whales (*Z. cavirostris*, *M. carlhubbsi*, *M. stejnegeri* and *M. densirostris*) are also considered to be social species, forming relatively small groups of 5-10 individuals. Mass strandings have occurred suggesting strong social bonds between individuals, however, information is very restricted as beaked whale species are rarely sighted (Mead 2002).

Beluga whale (*D. leucas*) and Narwhal (*M. monoceros*) form temporary groups of 5-10 individuals although large herds of up to 1000 individuals have been observed at times. Social structure is fluid, as short termed associations are formed between individuals within groups (Murray *et al.* 1999; Corry-Crowe 2002).

4.1.1.2 *Mysticeti* social behaviour

Associations between individual mysticetes are sometimes observed during migration, mating and foraging behaviour, but it is generally accepted that *Mysticeti* species tend to lead a much more solitary and asocial life than *Odontoceti* species (Bannister 2002; Connor 2002). Aggregations typically do not exhibit the high density and physical proximity seen for *Odontoceti*; individuals are usually dispersed over a

large area and rarely come into close contact (Valsecchi *et al.* 2002). In the *Mysticeti* species, such as the minke whales (*Balaenoptera acutorostrata* and *B. bonaerensis*), fin whale (*Balaenoptera physalus*), blue whale (*Balaenoptera musculus*), sei whale (*Balaenoptera borealis*), bowhead whale (*Balaena mysticetus*) and bryde's whale (*Balaenopetra edeni*) individuals have been observed to associate during reproduction and migration, although observational studies suggest that, on average, these associations are short termed (maximum of 1-2 days; Bannister 2002).

The exception is the Humpback whale (*Megaloptera novaeangliae*; Clapham 1996). Social behaviour has been observed on humpback whale feeding grounds, where groups of individuals have been observed to co-ordinate their movements and to co-operate with one another during feeding, coming into close physical proximity (Clapham 1996; Clapham 2002). These associations have been observed to last for 1-2 days to 1-2 weeks (Clapham 2002). In addition, males on the breeding grounds form groups to compete aggressively for access to females, by tail slashing, ramming and head butting (Tyack and Whitehead 1983). Furthermore, gray whale (*Eschrichtius robustus*) and right whales (*Eubalena glacialis*, *E. japonica* and *E. australis*) are also known to interact on breeding grounds (Bannister 2002).

4.1.2 Effects of social structure on disease transmission

Social structure has been shown to affect disease transmission and the emergence of epidemics (Murray 1993; Cliff 1995; White *et al.* 1996; Beardmore and White 2001). Studies of wild animal populations have shown a positive correlation between population density and pathogen abundance (Semple *et al.* 2001).

Studies in wild populations of foxes infected with rabies (*Rhabdoviridae*) have shown that there are two possible extremes. The first case is when viral replication occurs in the neocortex and the virus enters the spinal cord (Kaplan 1977; Bacon 1985). Infected foxes exhibit photophobia and paralysis, which results in the isolation of the individual and subsequent confinement to their den ('dump' rabies; Kaplan 1977; Bacon 1985). The second case is when the virus replicates in the limbic system (olfactory cortex, amygdala and hippocampus) and infected foxes lose all sense of direction and fear, which leads them to roam considerable distances and attack other animals ('furious' rabies; Kaplan 1977; Bacon 1985).

The effects of bovine tuberculosis (*M. bovis*) on wild populations of badgers have been extensively studied in Britain (Cheeseman and Mallison 1982). Infected badgers also exhibit two possible extremes. Extensive abscess formations in the lumbar region (lowest ribs and pelvis) lead to a very limited movement or paralysis. On the other hand, cases were observed where infected badgers displayed loss of orientation and fear leading infected individuals to roam randomly, intruding con-specific territories and engaging in territorial fights.

Cheeseman and Mallison (1983) and Murray (1993) have shown that, when infected individuals exhibited the symptoms which increase social behavior and physical contact, the rate of infection in the population increased considerably. In the aforementioned studies, the degree of interaction between groups was shown to positively correlate to infection rate (Cheeseman and Mallison 1983; Murray 1993).

The effects of two main forms of social behavior on disease dynamics have been studied in wild animal populations: a) non-interacting social groups and b) interacting social groups (Bowers *et al.* 1997; Beardmore *et al.* 2001). Beardmore *et al.* (2001) and Cudeldj *et al.* (2004) showed that in non-interacting social groups, survival of susceptible individuals arises because there is loss of infection (through the deaths of infected individuals) before infected and susceptible individuals (or groups) begin to interact. Furthermore, these studies showed that in the social groups which interact, the total number of susceptible individuals which survive the infection, is inversely proportional to the between-group contact rate.

4.1.3 Objectives

Disease dynamics are affected by the social organization of a given population and according to previous investigations, social species are under a higher pathogen pressure than less social species (Bowers *et al.* 1997; Beardmore *et al.* 2001; Cudeldj *et al.* 2004). In addition, cetacean species included in this study exhibit different forms of social behavior and differences in between-group contact rate. Chapter 3 showed that *DQB1* locus variation in cetaceans is under positive selection and involved in the immune response. The aims of this chapter are:

I) **To assess the effects of social behaviour in cetaceans on the evolution of the MHC *DQB1* locus**. Due to their well-developed social behaviour,

Odontoceti species are expected to be under stronger selection pressure than the less social *Mysticeti* species, as previous studies have shown that pathogen densities and transmission in social species is higher than in asocial species (Beardmore *et al.* 2001). This will be assessed through comparison at the sub-order level.

II) **To use the comparative analysis of different species to assess the effects of social structure on the evolution of *DQB1* locus**. Four main forms of social structure exist in cetaceans: a) Stable social structure based on kin relationships and small groups (*e.g.* the killer whale), b) unstable social structure based on short-term associations and large groups (*e.g.* bottlenose dolphins, finless porpoise), c) restricted social structure based on temporary associations during foraging and migrating (*e.g.* humpback whale) and d) no evident social structure (*e.g.* minke whales). Previous studies have shown that the degree of interaction between individuals affects disease transmission and infection; as the higher the interaction is, the less the number of individuals which survive the infection (Beardmore *et al.* 2001). Cetacean species which are known to exhibit high between-group contact and fluid society structure are expected to be under stronger pathogenic pressure than species which exhibit less between-group interaction and stable social structure.

4.2. RESULTS

4.2.1 Phylogenetic analysis

One of the main characteristics of MHC evolution is trans-species polymorphism where ancestral polymorphism is retained in the descendant species (see Chapter 1: 1.2.2; Klein 1980; Takahata 1992; Klein *et al.* 1998). The result is that MHC alleles from one species may be more similar to alleles from different species than to other alleles from the same species. This pattern of evolution has been observed in several species of vertebrates and it is suggested to be the result of retention of allelic lineages by balancing selection (Hughes and Nei 1988, 1989; Wills 1991; Takahata

1992; Klein *et al.* 1998; Nei *et al.* 2000). According to the expectations of trans-species evolution, alleles which are under higher selection pressure will be preserved for longer evolutionary periods (Hughes and Nei 1988, 1989). It was shown in Chapter 3 that the evolution of *DQB1* in cetaceans exhibits the trans-species evolution pattern. In this study phylogenetic analysis is used in order to assess differences in the persistence of *DQB1* alleles in the two cetacean suborders.

4.2.1.1 Phylogenetic analysis at the suborder level

The number of *Odontoceti* species used in this study represents 30% of extant species and the number of the *Mysticeti* species included in this study represents 40% of the extant mysticeti species, however, there is a higher total number of alleles present in *Odontoceti* species. Figures 4.2.1 and 4.2.2 show the Neighbor-Joining tree phylogeny of exon 2 *DQB1* alleles in *Odontoceti* and *Mysticeti* species respectively. The proportion of lineages that exhibit trans-species evolution was 57% in *Odontoceti* and 13% in *Mysticeti* phylogeny. There was a significant difference between the proportions of lineages showing trans-species evolution in *Odontoceti* and *Mysticeti* suborders (chi-square value=8.61, p-value=0.003, df=1).

The trans-species evolution pattern is also observed in the network phylogenies constructed by Median-Joining in both the *Odontoceti* and the *Mysticeti* phylogenies (Figures 4.2.3 and 4.2.4 respectively). The network method phylogeny displays alternative evolutionary paths in the form of circles and median vectors denote the possible extant un-sampled alleles or extinct ancestral alleles (see Chapter 2: 2.6.2). The mutation sites correspond to Figure 3.2.1 consensus numbering. The divergence of the alleles in the *Odontoceti* and *Mysticeti* species is mainly due to mutations, which are located in the β-sheet and α-helix portions of the protein.

Figure 4.2.1: Neighbor-Joining tree of *DQB1* exon2 sequences of all the *Odontoceti* species included in this study. Tree is based upon 1000 bootstrap replications. Numbers in bold font style above branches indicate bootstrap value, numbers below in regular style the distances. An asterisk denotes common alleles between *T. aduncus* and *T. truncatus* and between *B. bonaerensis* and *B. acutorostrata*.

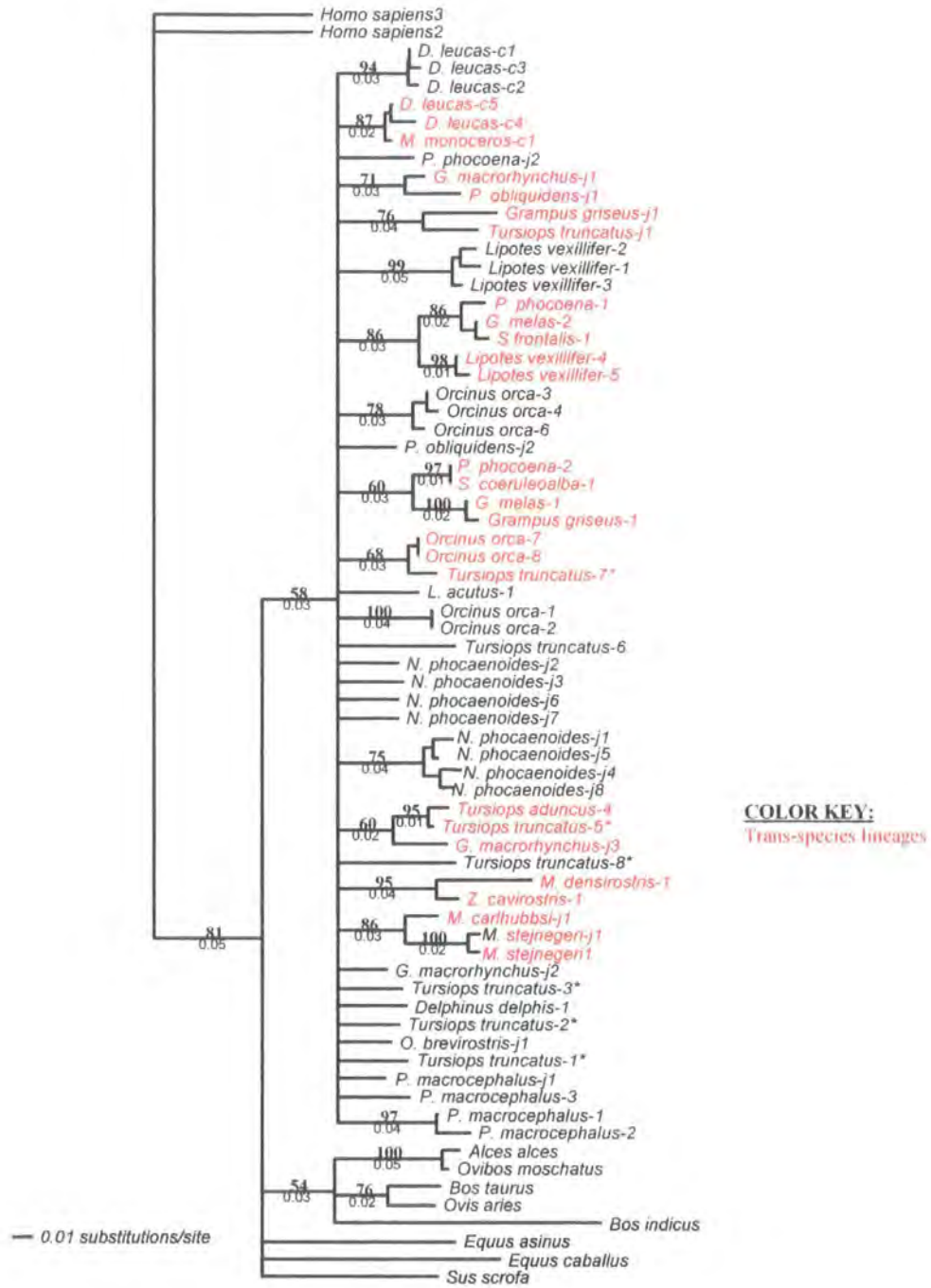


Figure 4.2.2: Neighbor-Joining tree of *DQB1* exon2 sequences of all the *Mysticeti* species included in this study. Tree is based upon 1000 bootstrap replications. Numbers in bold font style above branches indicate bootstrap value, numbers below in regular style the distances. An asterisk denotes common alleles between *T. aduncus* and *T. truncatus* and between *B. bonaerensis* and *B. acutorostrata*.

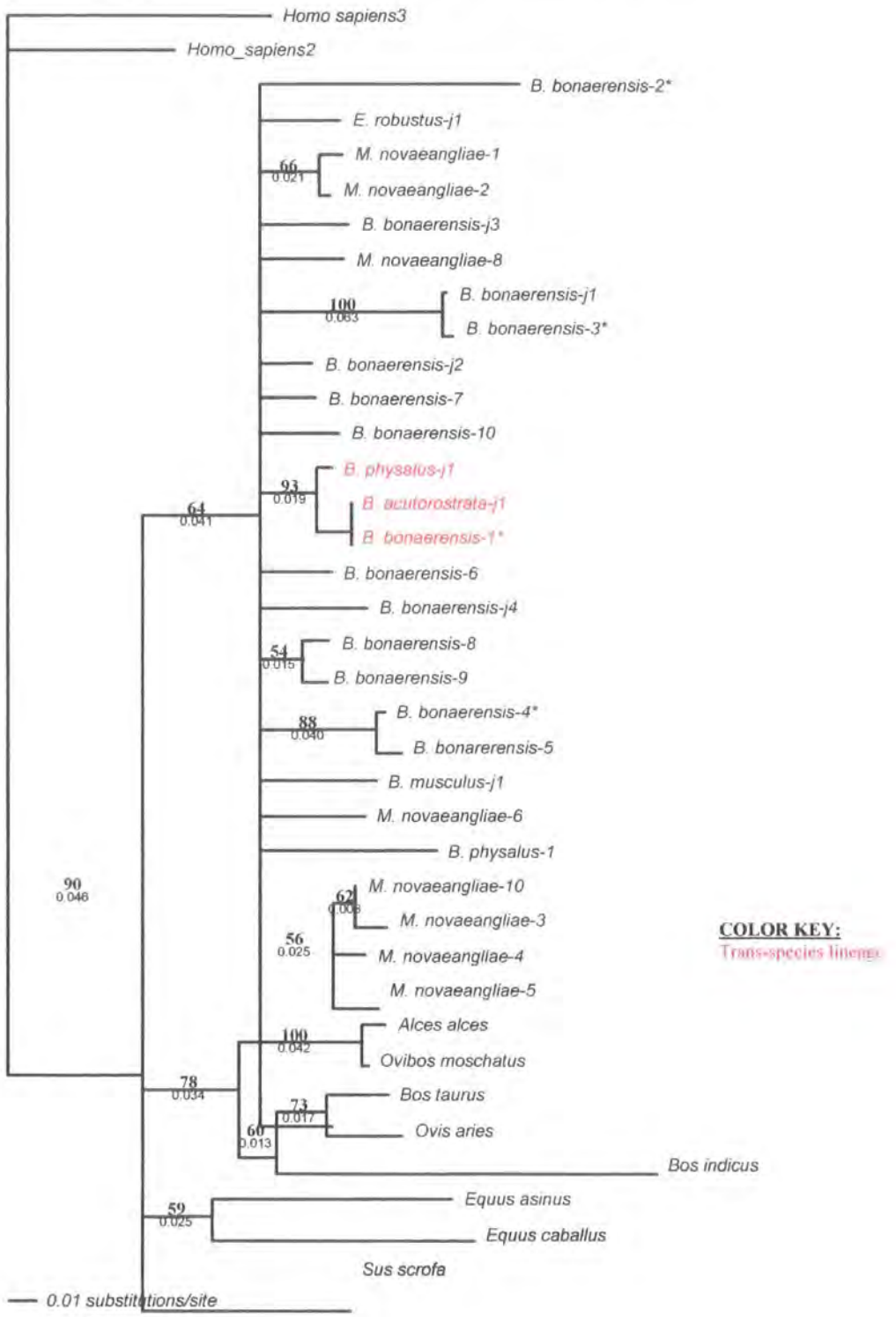


Figure 4.2.3: Median joining network of *Odontoceti DQB1* alleles included in this study.

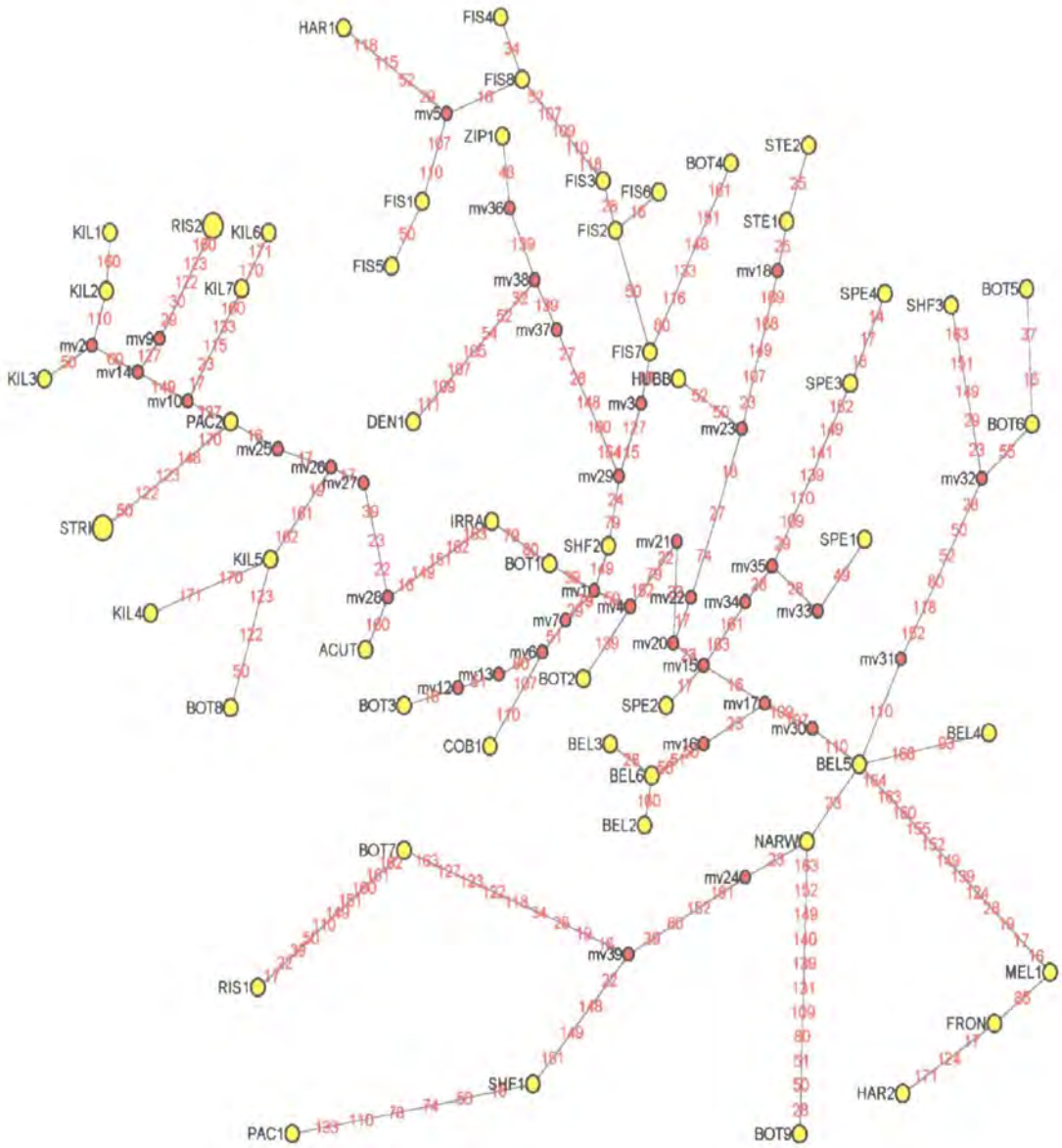
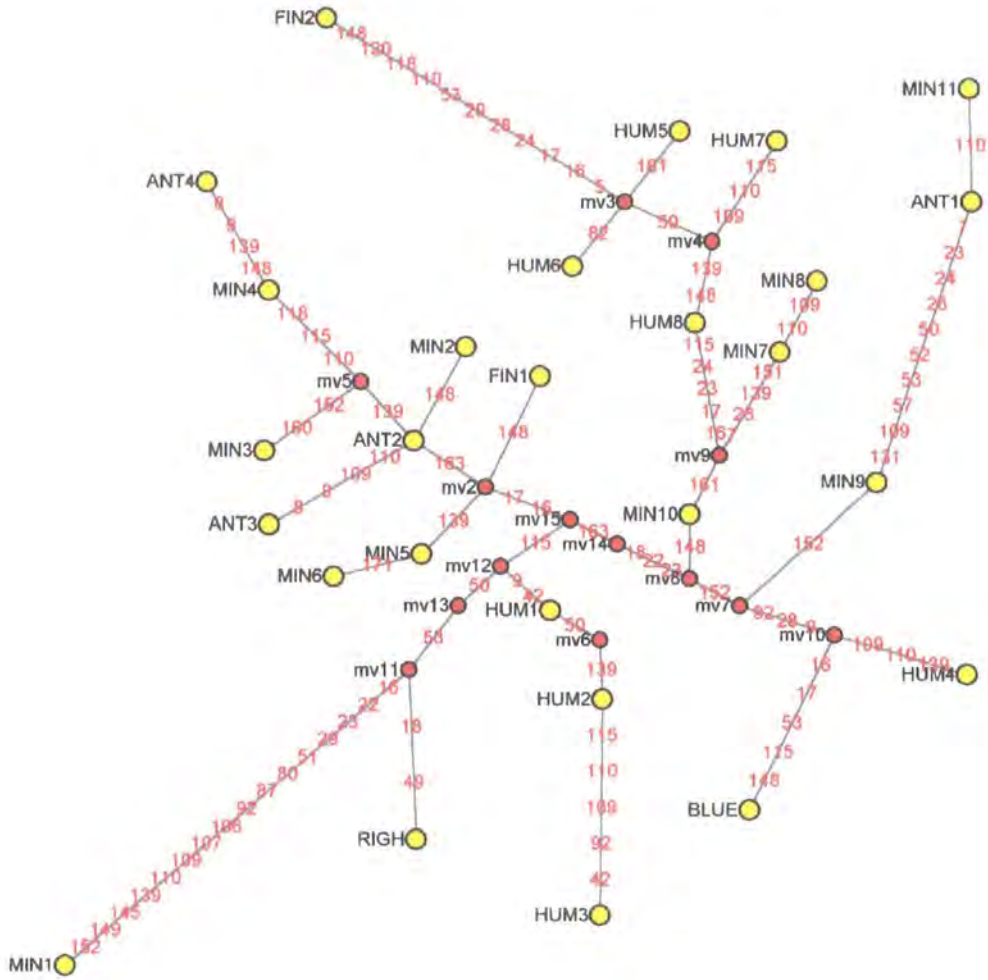


Figure 4.2.4: Median joining network of *Mysticeti DQB1* alleles included in this study.



4.2.1.2 Phylogenetic analysis at the species level

When polymorphic alleles in the MHC are maintained in a population by selection for a long time, allelic divergence occurs by accumulation of mutations (Takahata and Nei 1990). The number of median vectors and mutation steps is proportional to time; the longer alleles are maintained by selection, the higher the number of accumulated mutations (Takahata and Nei 1990).

Network phylogenies were constructed for *N. phocaenidae* (Figure 4.2.5), *D. leucas* (Figure 4.2.6), *T. aduncus* (Figure 4.2.7), *T. truncatus* (Figure 4.2.8), *O. orca* (4.2.9), *M. novaeangliae* (4.2.10), *B. acutorostrata* (4.2.11) and *B. bonaerensis* (4.2.12).

Table 4.2.1 shows the median vector number and average number of mutation steps between alleles (see Chapter 2). There is a possible correlation between the number of median vectors present and the ratio of non-synonymous to synonymous substitutions (ω) in PBR (Spearman's r = 0.912, p=0.011) and a negative correlation between the number of median vectors present and the rate of synonymous substitutions per synonymous site (*ds*; Spearman's r = -0.853, p=0.031).

Table 4.2.1: Shows the number of Median vectors (MV), average mutation number (μ) estimated by the network phylogeny analysis and *dn/ds* (ω) ratio in the PBR sites.

	MV	μ	ω
N. phocaenidae	2	1.300	NA
T. aduncus	7	3.417	31.738
T. truncatus	7	3.429	29.282
D. leucas	0	2.200	NA
O. orca	2	2.444	2.702
B. acutorostrata	5	3.636	11.241
B. bonaerensis	5	4.000	12.667
M. novaeangliae	3	2.727	2.047

4.2.2 Substitution rate analysis

Hughes and Nei (1988, 1989) have demonstrated that balancing selection enhance the rate of non-synonymous substitutions in MHC. Codons which are under positive selection exhibit high rates of non-synonymous over synonymous substitutions whilst codons which are not under positive selection exhibit high rates of synonymous over non-synonymous substitutions (Lopez de castro *et al.* 1982; Hughes and Nei, 1988, 1989; Wills 1991). In Chapter 3 (Table 3.2.2) it was shown that *DQB1* in cetaceans exhibits high rates of non-synonymous substitutions suggesting that it is under positive selection. Substitution analysis in this study is used in order to assess differences in substitution rates between *Odontoceti* and *Mysticeti DQB1* alleles.

The approximate method of Nei-Gojobori, (1986) and the Maximum Likelihood method of Yang *et al.* (2000) were used in order to estimate the non- synonymous and synonymous substitution ratio (ω) in *DQB1* exon2. Ratios were estimated for the full

sequence, the Peptide Binding Region (PBR) and the non Peptide Binding Region (nonPBR) of exon-2 for the *Odontoceti* and *Mysticeti* species.

Figure 4.2.5: Median-joining network of *P. phocaenoides DQB1* alleles identified in the study by Hayashi *et al.* (2004).

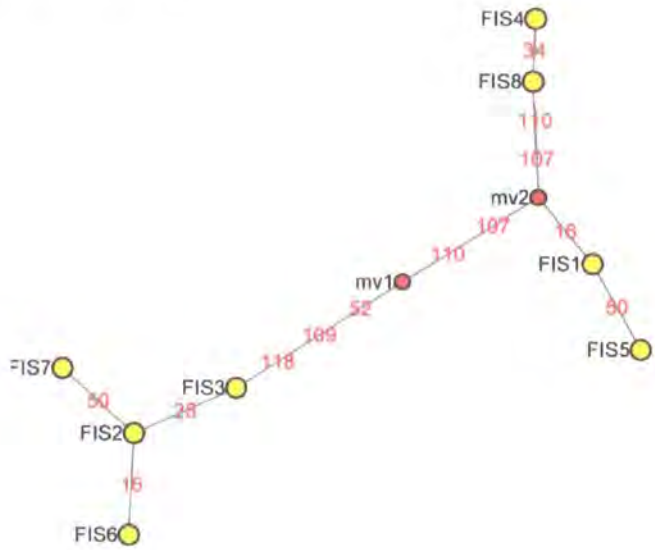


Figure 4.2.6: Median-joining network of *D. leucas DQB1* alleles identified in the study by Murray *et al.* (1995).

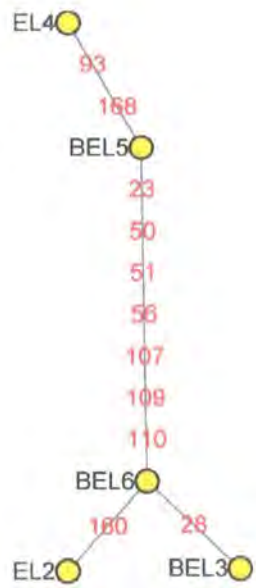


Figure 4.2.7: Median-joining network of *T. aduncus DQB1* alleles identified in this study.

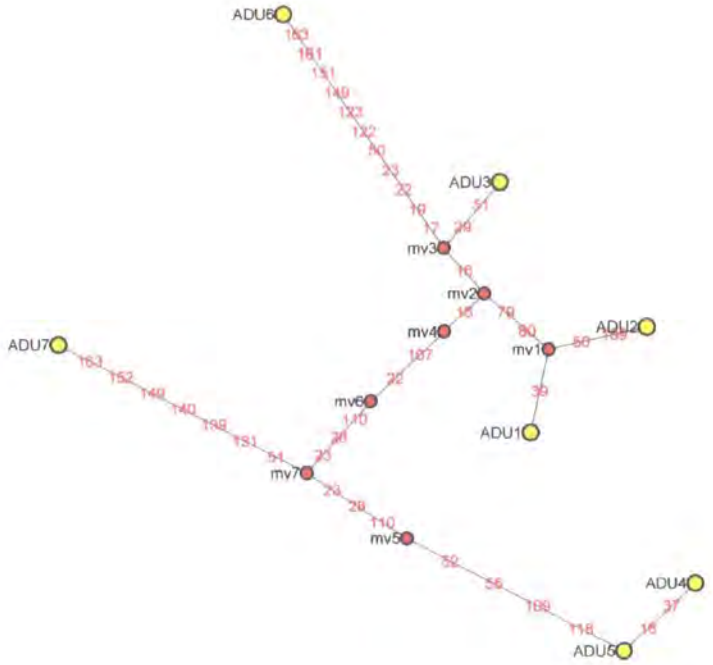


Figure 4.2.8: Median-joining network of *T. truncatus DQB1* alleles identified in this study.

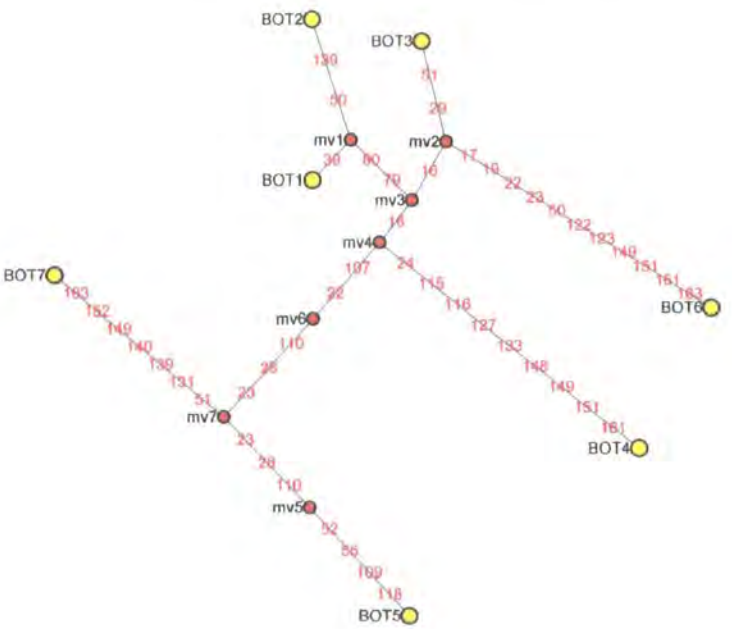


Figure 4.2.9: Median-joining network of *O. orca DQB1* alleles identified in this study.

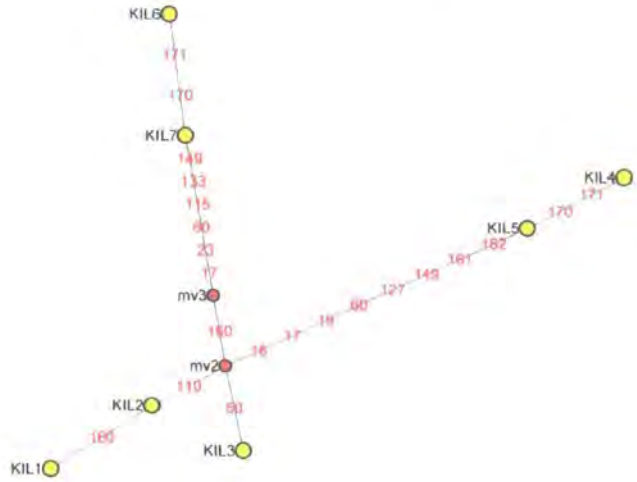


Figure 4.2.10: Median-joining network of *M. novaeangliae DQB1* alleles identified in this study.

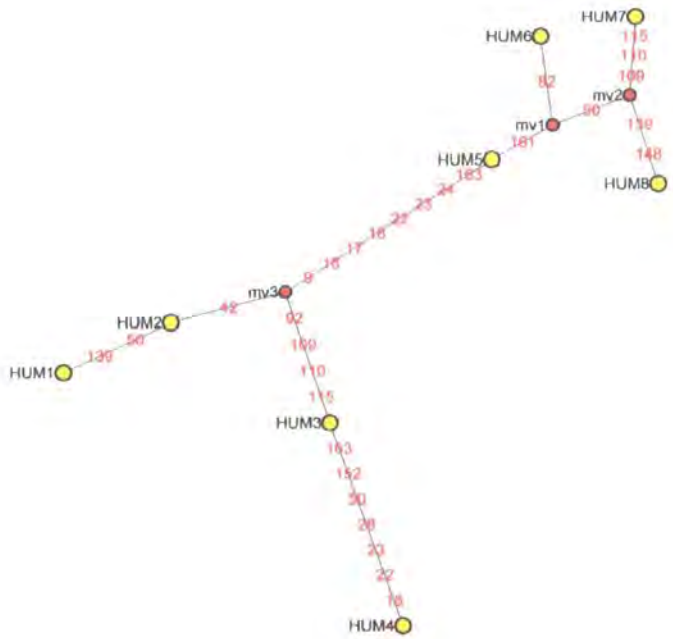


Figure 4.2.11: Median-joining network of *B.acutorostrata DQB1* alleles identified in this study.

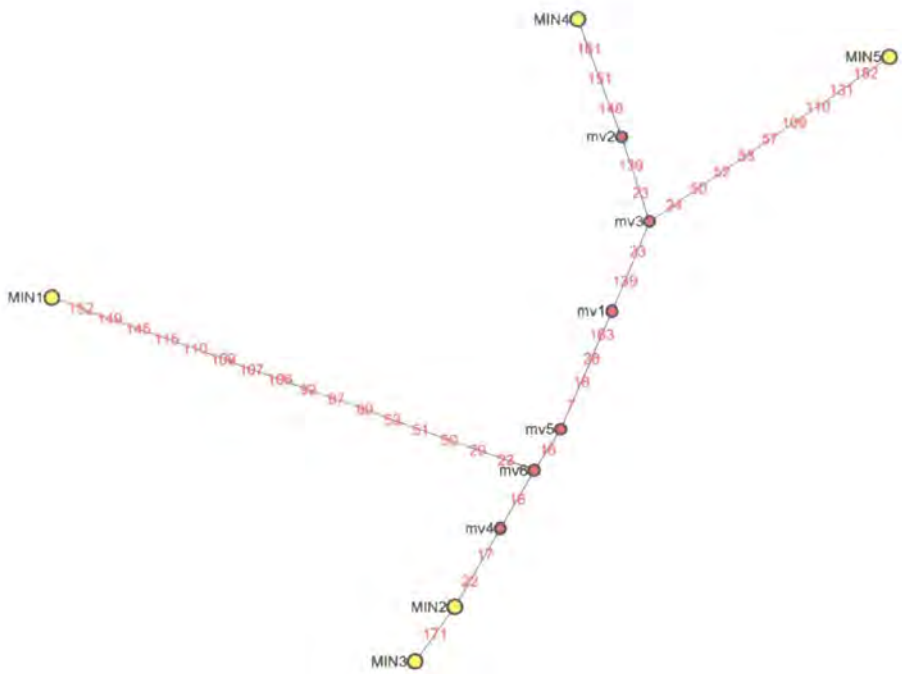
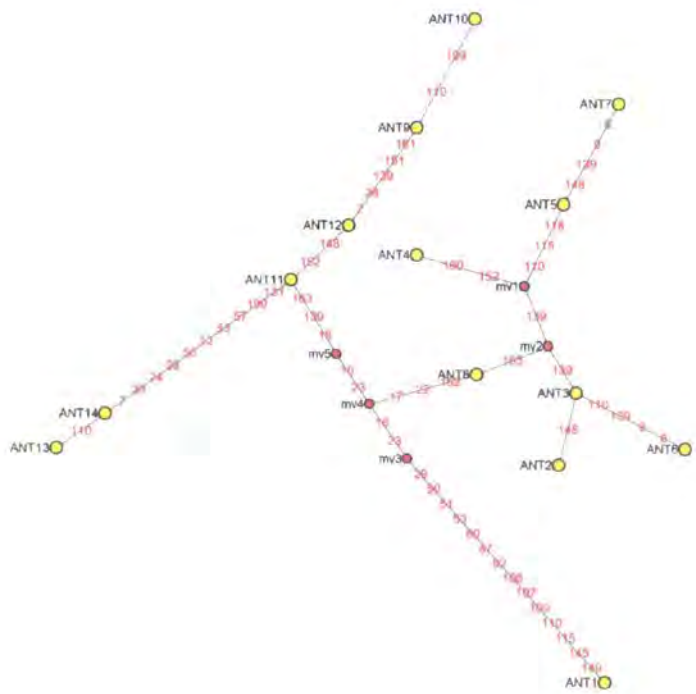


Figure 4.2.12: Median-joining network of *B. bonaerensis DQB1* alleles identified in this study.



The group of six *Odontoceti* species (*O. orca*, *D. leucas*, *M. monoceros*, *P. macrocephalus*, *Z. cavirostris* and *M. steijeneri*) denoted as *Odontoceti* (low) exhibit the lowest ω ratios among *Odontoceti.* This group was used in order to determine whether observed differences between mysticetes and odontocetes are due to high ratios at the species level rather than at suborder level, as the *Odontoceti* sample size is twice as big as the *Mysticeti* sample size. However, the t-test pairwise comparison shows that there is no significant difference between any of the groups; t-values range from 0.03 to 1.54 and p-values from 0.137 to 0.941 and degrees of freedom (df) were 112 for Full length, 26 for PBR and 84 for non-PBR.

Table 4.2.2: Number of synonymous substitutions per synonymous site (*ds*) and number of non-synonymous substitutions per non synonymous site (*dn*) in the A) full length of the exon-2 sequence, B) putative non-PBR and C) putative PBR sites respectively, of all the species included in this study. Numbers in brackets indicate Standard Error and ω is the ratio of non-synonymous to synonymous substitutions.

	FULL			PBR			non-PBR		
	dn	ds	ω	dn	ds	ω	dn	ds	ω
Nei-Gojobori									
Odontoceti	0.1168 (0.0253)	0.0204 (0.0069)	**5.725**	0.3238 (0.1976)	0.0197 (0.0105)	**16.437**	0.0625 (0.0200)	0.0536 (0.0212)	**1.166**
Mysticeti	0.0844 (0.0233)	0.0381 (0.0160)	**2.215**	0.1976 (0.0651)	0.0157 (0.0159)	**12.586**	0.0536 (0.0212)	0.0440 (0.0210)	**1.218**
Odontoceti (low)	0.1073 (0.0242)	0.0309 (0.0106)	**3.472**	0.2609 (0.0807)	0.0211 (0.0122)	**12.365**	0.0662 (0.0217)	0.0330 (0.0135)	**2.006**
ML (Kosakovsky *et al.*2004)									
Odontoceti	1.692	0.680	**2.489**	3.542	0.232	**15.277**	0.846	1.079	**0.784**
Mysticeti	0.848	0.619	**1.370**	1.468	0.147	**10.021**	0.448	0.679	**0.660**
Odontoceti (low)	0.494	0.238	**2.068**	1.272	0.132	**9.638**	0.407	0.576	**0.707**

Substitution analysis was also performed at species level, the ω ratios of *T. truncatus*, *T aduncus*, *O. orca*, *D. leucas*, *N. phocaenidae*, *B. acutorostrata*, *B. bonaerensis* and *M. novaeangliae* were estimated (Table 4.2.3). There was a significant difference (after Bonferonni correction α=0.01, k=6) at the PBR between *T. aduncus* and *M. novaeangliae* (t-value=3.03, p-value=0.005, df=26), *T. aduncus* and *O. orca* (t-value=3.73, p-value=0.001, df=26), *T. truncatus* and *M. novaeangliae* (t-value= -2.25, p-value=0.033, df=26), *T. truncatus* and *O. orca* (t-value=-3.68, p-value=0.001, df=26), *B. bonaerensis* and *M. novaeangliae* (t-value=-2.22, p-value=0.035, df=26), *B.*

bonaerensis and *O. orca* (t-value=-3.60, p-value=0.001, df=26), *B. acutorostrata* and *M. novaeangliae* (t-value=2.34, p-value=0.021, df=26) and *B. acutorostrata* and *O.orca* (t-value=-3.01, p-value=0.006, df=26). In the full-length comparison, there was no significant difference with t-values ranging from 0.03 to 2.08 and p-values from 0.062 to 0.976 and df=112. In addition, in the non-PBR sites comparison there was also no significant difference with t-values ranging from 0.12 to 0.98 and p-values from 0.329 to 0.905.

Table 4.2.3: Number of synonymous substitutions per synonymous site (*ds*) and number of non-synonymous substitutions per non synonymous site (*dn*) in the A) full length of the exon-2 sequence, B) putative non-PBR and C) putative PBR sites respectively, of the eight focal species. Numbers in brackets indicate Standard Error and ω is the ratio of non-synonymous to synonymous substitutions.

	FULL			PBR			non-PBR		
	dn	ds	ω	dn	ds	ω	dn	ds	ω
Nei-Gojobori									
O. orca	0.0578 (0.0184)	0.0254 (0.0162)	**2.276**	0.1078 (0.0520)	0.0297 (0.0317)	**3.630**	0.0428 (0.0185)	0.0233 (0.0184)	**1.837**
M. novaeangliae	0.0596 (0.021)	0.0536 (0.027)	**0.899**	0.0974 (0.0589)	0.0296 (0.0329)	**3.291**	0.0432 (0.0185)	0.0648 (0.0330)	**0.667**
N. phocaenoides	0.0401 (0.0151)	0	**NA**	0.0989 (0.0389)	0	**NA**	0.0204 (0.0160)	0	**NA**
B. acutorostrata	0.1176 (0.0296)	0.0599 (0.0253)	**1.963**	0.3318 (0.1135)	0.0148 (0.0165)	**22.419**	0.0626 (0.0253)	0.0733 (0.0327)	**0.854**
B. bonaerensis	0.0870 (0.0232)	0.0339 (0.0170)	**2.566**	0.2451 (0.0753)	0.0087 (0.0062)	**28.173**	0.0517 (0.0220)	0.0402 (0.0216)	**1.286**
T. truncatus	0.1122 (0.0247)	0.0086 (0.0068)	**13.047**	0.3309 (0.0768)	0.0089 (0.0071)	**37.180**	0.0535 (0.0193)	0.0083 (0.0087)	**6.446**
T. aduncus	0.1050 (0.0243)	0.0079 (0.0069)	**13.292**	0.2992 (0.0709)	0.0050 (0.0062)	**59.840**	0.0519 (0.0204)	0.0088 (0.0098)	**5.898**
D. leucas	0.0440 (0.0171)	0.0093 (0.0090)	**4.731**	0.1655 (0.0612)	0	**NA**	0.0244 (0.0158)	0.0111 (0.0110)	**2.198**
ML (Kosakovsky et al)									
O. orca	1.683	0.951	**1.599**	2.702	0.843	**2.702**	1.014	0.985	**1.030**
M. novaeangliae	0.927	0.790	**1.173**	1.519	0.742	**2.047**	0.365	0.675	**0.541**
N. phocaenoides	1.294	0	**NA**	3.634	0	**NA**	2.987	0	**NA**
B. acutorostrata	0.837	0.621	**1.350**	3.833	0.341	**11.241**	0.428	0.781	**0.549**
B. bonaerensis	0.421	0.279	**1.509**	2.711	0.214	**12.667**	0.724	0.564	**1.284**
T. truncatus	4.049	0.429	**9.438**	4.890	0.167	**29.282**	1.593	0.452	**3.525**
T. aduncus	5.003	0.624	**8.018**	3.618	0.114	**31.738**	0.975	0.343	**2.841**
D. leucas	1.403	1.254	**2.119**	2.560	0	**NA**	0.462	0.574	**0.805**

4.3 Discussion

In chapter 3 it was shown that *DQB1* exon 2 in cetaceans is under positive selection. This was illustrated through trans-species evolutionary patterns (allelic persistence) and substitution analysis (Figures 3.2.1, 3.2.2 and 3.2.3 and Table 3.2.2 and 3.2.3). In this chapter, *DQB1* evolution in cetaceans is taken one step further by investigating the pattern of evolution in the two cetacean sub-orders, *Odontoceti* and *Mysticeti*. The main aim of the present study is to assess whether social behaviour differences affect selection pressure and the pattern of evolution of MHC *DQB1* in cetacean species. Previous studies have shown that social species with high population densities and high rates of physical contact are under higher pathogenic pressures than less social species which exhibit low population densities and low rates of physical contact (Bowers *et al.* 1997; Beardmore *et al.* 2001; Read and Keeling 2003; Gudelj and White 2004).

According to the first hypothesis, *Odontoceti* species, which exhibit well-developed social behavior, will be under higher selection pressure than the less social *Mysticeti* species. Support was found for the first hypothesis, as the proportion of trans-species lineages present in the *Odontoceti* Neighbor-Joining phylogeny was significantly higher than the proportion of trans-species lineages exhibited by the *Mysticeti* species (Figures 4.2.1 and 4.2.2). However, nucleotide substitution analysis demonstrated that there was no significant difference between the ω ratios of *Odontoceti* and *Mysticeti* species at the suborder level.

For the purpose of this study, social structure in cetaceans has been divided into four broad categories in order to assess the effect of between-group interaction (social behavior) and social structure stability at the species level (see 4.1.3). According to the second hypothesis, pathogenic pressures are proportional to the degree of physical interaction and unstable social structure, which hypothetically may affect the evolution of *DQB1* in cetaceans (Gudelj and White 2004). Support was found for the second hypothesis.

The network analysis has shown that there is a positive correlation between the number of median vectors and ω ratios, an observation which is consistent to the expectations of overdominant selection (see Chapter 6 for a detailed discussion). Killer whale and humpback whale exhibit the lowest median vector number and also the lowest ω value in the PBR sites. Furthermore, the results drawn from the sequence

analysis from both the approximate method of Nei-Gojobori (1986) and Maximum Likelihood method of Yang *et al.* (2000) indicate that there was a significant difference in the species comparison (Table 4.2.5). *T. truncatus* and *T. aduncus* show significantly higher ω ratios in full length of the sequence and the Peptide Binding Region (PBR) of exon-2. This observation supports the hypothesis that unstable social structure based on short termed associations and large groups (*T. truncatus* and *T.* aduncus) may be under higher pathogenic pressure than a stable social structure based on kin relationships and small groups (*i.e.* killer whale).

Nevertheless, there is no significant difference between *M. novaeangliae* and *O. orca* and both minke whales *B. acutorostrata* and *B. bonaerensis* have significantly higher ω ratio in the PBR than *M. novaeangliae* and *O. orca*. These observations do not conform to the expectations of the hypothesis; according to previous studies within these species, the most social is *O. orca* while *M. novaeangliae* has been suggested to show the most developed social behaviour among *Mysticeti* species (Clapham 2002; Ford 2002).

Body size (Semple *et al.*, 2002) promiscuity (Nunn *et al.* 2000; Anderson *et al.* 2003) and life span (Van Boven *et al.* 2003) have been suggested to play a key role in disease dynamics (Hempel 2003; Gudelj 2004). For example, it has been suggested that in primates large body sizes provide relatively lower surface area to infections than smaller body sizes (Semple *et al.* 2001). In addition, there is a positive correlation between higher counts of white blood cell and red blood cell and occurrences of multiple partner mating, an observation which is attributed to higher selection pressure in promiscuous species (Anderson *et al.* 2003). Furthermore, it has also been suggested that it may be more cost-effective for a species with a short life span to invest in the immune response than a species with a longer life span (Van Boven and Weissing 2004).

Differences in body size and life span are also observed in cetacean species. For example, finless porpoise is among the smallest of *Odontoceti* species with an average length of 1.5m and average weight of 40kg whilst killer whale is among the largest *Odontoceti* with an average length of 9 meters and average weight of 6 tonnes (Carwardine 1995). *Mysticeti* species also exhibit a difference in size, for example *B. acutotorostrata* is among the smallest of *Mysticeti* species with an average length of 9 meters and average weight of 7 tonnes whilst humpback whale is among the largest with an average length of 14m and average weight of 30 tonnes (Clapman 2002; Valsecchi *et*

al. 2002). In addition, it has been observed in cetaceans that there is a correlation between body size and life span (Chivers 2002). The approximate average life-span for killer whale is 75 years (Ford 2002), for bottlenose dolphins is 30 years (Scott and Wells 2002), for finless porpoise is 20 years (Read 2002), for humpback is 70 years (Clapham 2002) and for the minke whales is 40 years (Perrin and Brownell 2002). Life span and body size have been suggested to affect the evolution of the immune response (Lochmiller and Deerenberg 2000; Semple *et al.* 2001; Zuk and Stoehr 2002; Van Boven and Weissing 2004). The results of the present study hypothetically may suggest that life span and body size may affect the evolution pattern of the *DQB1* locus. The significantly low non-synonymous substitution rates in killer whale and humpback whale in the PBR sites, may suggest that other life history parameters (such as longevity and body size) besides social behavior and social structure may affect the evolution of *DQB1* locus in cetaceans.

In the majority of *Mysticeti* species, life history is divided into three phases: feeding in high latitudes, mating and calving in low latitudes and an annual migration between the feeding grounds and the breeding grounds (Bannister 2002). In the present study, humpback whale (*M. novaeangliae*) samples were taken from the North Pacific feeding grounds. North Atlantic humpback whales migrate between the feeding grounds in Alaska, California, Bering Sea and western North Pacific and the breeding grounds in Hawaii, Mexico and Japan (Clapham 2002). The feeding grounds of the common minke whale included in this study (*B. acutorostrata*), extend to Baffin bay, Denmark strait and Barents sea while the breeding grounds extend to the Caribbean, Straits of Gibralter and North West Africa (Senegal) and in the North Atlantic (Perrin and Brownell 2002). In the North Pacific, feeding grounds extend to Bering Sea and Chuksi Sea and breeding grounds extend to Mexico, Hawaii and Indonesia (Perrin and Brownell 2002). The feeding grounds of the Antarctic minke whale (*B. bonaerensis*) extend to the ice edge in Antarctic and the breeding grounds in the northeastern and eastern Australia, western South Africa and northeastern coast of Brazil (Perrin and Brownell 2002). The North pacific common minke whale is seasonally sympatric to the Antarctic minke whale in the feeding grounds (Perrin and Brownell 2002).

Studies on the spread of raccoon rabies virus in eastern USA have shown that animal dispersal can have a profound effect on disease dynamics, as infected individuals spread disease away from the sites of original introduction and disease is maintained in the population for longer periods (Wilson *et al.* 1997; Russel *et al.* 2003). The spatially

explicit model by Gudelj *et al.* (2004) and Gudelj and White (2004) suggests that localized infection decreases the number of susceptible individuals locally and hence the levels of infection, while animal dispersal spreads disease throughout the population and increases the number of susceptible individuals and infection levels. In addition, Semple *et al.* (2002) have shown that primate species which live in habitats and which exhibit different environmental conditions (such as temperature) also exhibit differences in white blood cells (leucocytes, lymphocytes and platelets). This study suggested evolutionary adaptation of the immune response in primates to ecological factors, as different habitats also exhibit differences in pathogenic pressures (Stuart and Strier 1995; Sorci *et al.* 1997; Semple *et al.* 2002).

The two species of minke whales included in this study, exhibited a significantly higher ω ratio value in the PBR sites than humpback whale and killer whale. However, there is no evident social structure in minke whales and both body size and life span parameters are comparable to the killer whale (see above). Studies in terrestrial mammals have shown that animal dispersal and habitat differences may affect the pathogenic pressures exhibited by a given species (Semple *et al.* 2002; Russel *et al.* 2003; Gudelj and White 2004). The results of the present study may hypothetically suggest that minke whales experience higher pathogenic pressures due to their broad worldwide distribution.

4.3.1 Conclusions

In Chapter 3, it was shown that exon-2 of class II *DQB1* locus in cetaceans is under positive selection. The general aims of this chapter were: i) to evaluate *DQB1* evolution pattern at the suborder and species level and ii) to assess whether phylogenetic and nucleotide substitution analyses on this locus can provide further information on pathogenic pressures differences attributed to life history characteristics.

Trans-species evolution pattern exhibited by both sub-orders and high rates of non-synonymous substitutions both at the sub-order and at the species level provide further support to the hypothesis that this locus in cetaceans is under positive selection. In addition, the substitution analysis at the species level showed that substitution rates among species significantly differ, an observation which might be attributed to pathogenic pressure differences, through life history characteristics such as social structure and social behavior, body size, longevity and species distribution.

Previous studies in terrestrial mammals have shown that these parameters may affect pathogenic pressure and the evolution of the immune response (Bowers *et al.* 1997; Beardmore *et al.* 2001; Semple *et al.* 2002; Anderson *et al.* 2003; Read and Keeling 2003; Russel *et al.* 2003; Gudelj and White 2004). The results of the present study may suggest that the evolution of the *DQB1* locus in cetaceans may be affected by life history characteristics.

CHAPTER 5:

DQB1 VARIATION AND POPULATION GENETIC STRUCTURE IN CETACEANS

5.1 INTRODUCTION AND OBJECTIVES

5.1.1 Population genetics of the six focal cetacean species

5.1.1.1 Population genetics of the Killer whale (*Orcinus orca*)

Killer whale populations can be found across all major oceans in both polar and temperate waters and in particular in coastal areas of high productivity (Dahlheim and Heyning 1999; Ford 2002; Hoelzel *et al.* 2002). The social associations formed by this species are very stable and there are regional populations with identified individuals which are known to have persisted for decades (Ford *et al.* 1998; Ford 2002; Hoelzel *et al.* 2002). Sympatric populations found in the northeast Pacific differ extensively in ecology, genetic structure, behaviour and distribution patterns (Hoelzel and Dover 1991; Hoelzel *et al.* 1998; Ford and Ellis 1999; Hoelzel *et al.* 2002).

Initially, populations in the northeast Pacific were defined according to behavioral and geographical distribution and this resulted in the designation of at least three distinct populations: the 'transient' population, which prey on marine mammals, the 'resident', which prey on fish and the 'offshores' (Bigg *et al.* 1990; Dahlheim and Heyning 1999; Ford *et al.* 2000). The distribution of these populations appears to depend on their feeding habits. Thus residents show a strong seasonal movement associated with salmon migration while transients show only a weak seasonal pattern, probably due to the fact that their prey is abundant all year round in coastal waters. Unfortunately, little is known regarding the dietary and seasonal movement behavior of the offshore population (Ford 2002).

Killer whales show a dietary specialization among sympatric populations and each of the foraging techniques include a high level of skills and tactic accomplished through trial and error of previous individuals and successful strategies passed across the next generation (Ford 2002). The dietary specialization exhibited by the sympatric populations in the northeastern Pacific may have resulted to the effective isolation of the transients and the residents both socially and reproductively (Ford 2002). It has been shown that the resident, the transient and the offshore populations are all genetically distinct based on mitochondrial and microsatellite DNA analysis with the offshore being more closely related to residents than to transients (Hoelzel *et al.* 1998; Barrett-Lenard 2000; Hoelzel *et al.* 2002). Hence, in the northeastern Pacific, genetic analyses have identified at least six populations: Alaskan transients (SEAT), Southeastern Alaska residents (SEAR), Southern residents (SR), Northern residents (NR), Californian transients (CAL) and Offshore. There are also two resident populations which have been identified in the Berring Sea (BER) in North Western Pacific and in Iceland, in the North Eastern Atlantic (ICE; Hoelzel *et al.* unpublished data).

5.1.1.2. Population genetics of the Bottlenose dolphins *Tursiops truncatus* and *Tursiops aduncus*

The bottlenose dolphin is found in all the major oceans, from cold temperate to tropical seas, in coastal and offshore waters (Shane *et al.* 1996; Wells and Scott 2002). *T. truncatus* exhibits habitat differentiation among populations across its range, as sympatric or parapatric populations will use the coastal (nearshore) or the pelagic (offshore) environment (Hoelzel *et al.* 1998; Hoelzel *et al.* 2002; Natoli *et al.* 2004).

Studies on mtDNA and nuclear DNA (microsatellites) have shown that coastal and pelagic *T. truncatus* populations in the western North Atlantic are genetically differentiated (Hoelzel *et al.* 1998). In addition, in South Indian and South Pacific coastal habitats, a smaller morphotype of *T. truncatus* has been described as *T. aduncus* (or 'aduncus' form). A study on mtDNA by Wang *et al.* (1999) demonstrated that the *T. aduncus* coastal populations in China are genetically differentiated from *T. truncatus*, supporting the hypothesis that *T. aduncus* is a different species.

Previous population genetic analyses on mtDNA and microsatellites have shown that there is genetic differentiation between *T. truncatus* coastal populations from Mediterranean Sea (MS), South Africa (SAT), Eastern North Atlantic (ENA), Western North Atlantic pelagic (WNAP) and Western North Atlantic coastal (WNAC) and *T.*

aduncus populations from South Africa (SAA) and Southern China Sea (Hoelzel *et al.* 1998; Natoli *et al.* 2004). These studies suggested a high level of genetic differentiation among regional populations and showed that geographic regions and habitat use may affect the population structure in the *Tursiops* genus.

5.1.1.3 Population genetics of the Minke whales *Balaenoptera acutorostrata* and *Balaenoptera bonaerensis*

Minke whales have a worldwide distribution although temporal; observational studies suggest that their seasonal migration patterns are between breeding grounds in temperate waters and feeding grounds in polar waters in local winter and summer respectively (Brownell *et al.* 2000; Perrin and Brownell 2002). However, there have been observations of individuals in the Antarctic which did not exhibit seasonal migration (Perrin and Brownell 2002). The Antarctic Minke whale, *B. bonaerensis*, is found in the southern hemisphere and the Northern Minke whale, *B. acutorostrata*, in the Northern hemisphere (Brownell *et al.* 2000). Their seasonal migration patterns keep these two main populations apart (Brownell *et al.* 2000).

Hoelzel and Dover (1991) studied mtDNA variation in minke whale populations from West Greenland, western North Pacific and Antarctic management areas 4 and 5. Results of this study suggested that the genetic distance between Antarctic areas IV and V was very low, while the genetic distance between West Greenland and North Pacific was very high. In addition, high genetic distance was also observed between the comparisons of both northern populations to the Antarctic populations. Furthermore, a study by Wada (1991) suggested that the southern hemisphere minke whales (referred to as *B. bonaerensis*) should be separated from the North Pacific population (referred to as *B. acutorostrata*).

Van Pijlen *et al.* (1995) investigated the genetic structure of populations from the North Atlantic (NA), North Pacific (NP) and the Antarctic Oceans (management areas 4 and 5). The results from microsatellite and mitochondrial DNA analyses showed that the highest levels of genetic distance were observed between *B. acutorostrata* populations from the North Atlantic and the North Pacific and *B. bonaerensis* populations of Area 4 and 5 from the Antarctic (Van Pijlen *et al.* 1995). Moreover, the genetic divergence between the North Atlantic and North Pacific populations was shown to be significant while there was no significant difference between the Antarctic populations.

5.1.2 MHC variation and population genetics

By using MHC in population studies, one can compare population differentiation patterns between neutral markers and functional genes and thus, assessing the role of selection and neutrality on genetic diversity and also on MHC evolution. Boyce *et al.* (1996) examined the variation of *DQB1* and *DRB3* loci of class II MHC molecules and three microsatellite loci in bighorn sheep (*His canadensis*). It was shown that mean *Fst* values for both markers were not significantly different across populations of different geographic regions and therefore it was suggested that neutral forces such as genetic drift and gene flow influenced differentiation of MHC loci.

Landry and Bernatchez (2001) showed that in the Atlantic salmon (*Salmo salar*) differences in the environment and geographical scales correlate with significant differences in MHC allelic frequencies; however, overall F_{ST} values were not significant different from microsatellite loci values. They suggested that, although selection was evident in MHC variation in Atlantic salmon, the population structure inferred by MHC analysis was shaped mainly due to genetic drift and migration than selection (Klein 1987).

5.1.2.1 Functional basis of MHC alleles

An important application of MHC genetics to population studies is the mapping of allelic substitutions on the inferred structural model of the MHC molecule in order to determine differences in the actual function of the molecule itself according to local selective pressures (Hughes *et al.* 1996; Ou *et al.* 1998; Cohen 2002). For example, the study on the effects of pollution to MHC variation in estuarine fish showed that the population adapted to severe chemical pollution had specific amino acid substitutions in the a-helix portion (Cohen 2002). Furthermore, the fish from the unpolluted area also exhibited a significant different pattern in the b-pleated sheet of the PBR (Cohen 2002). Functional analysis has also been used in human studies, where human pathologies have been correlated to specific amino acid replacement and motif changes in the PBR of the MHC molecule among different populations (*e.g.* Nepom and Erlich 1991; Winchester 1994; Hill 1998; Ou *et al.* 1998;).

5.1.3 Objectives

Klein (1987) suggested that MHC alleles may experience periods of neutral evolution, during which genetic drift and mutation may be more prominent in maintaining MHC polymorphism than selection; however, selection occurs when populations change habitats and environmental conditions and thus new pathogens. Therefore, MHC is exposed to periods of neutral evolution which alternates with sporadic events of selection (Klein 1987). Previous studies in MHC population genetics on Scandinavian beavers (Ellegren *et al.* 1993), bighorn sheep (Boyce *et al.* 1996), Amerindians (Cerna *et al.* 1993; Valtes *et al.* 1999), Australian bush rat (Seddon and Baverstock 1999) and Atlantic salmon (Landry *et al.* 2001) have shown low levels of variation, suggesting that the role of selection in these populations may be weak (Cerna *et al.* 1993; Ellegren *et al.* 1993; Valtes *et al.* 1999).

Population genetic studies, using neutral markers on the six focal species included in this study, have shown that each species exhibits genetically differentiated populations. A comparison between populations which occupy different habitats may result in MHC genes being under different selection pressure, allowing the possibility to assess the role of selection in MHC (Stet and Egberts 1991). The main aims of this study are:

I) **To assess the role of selection in shaping variation of the *DQB1* locus in cetaceans.** Previous studies in terrestrial mammals compared the pattern of genetic structure at MHC and microsatellites and have shown that MHC variation may be affected by neutral evolution. It was shown that the *DQB1* locus in cetaceans is under positive selection at the order, suborder and species level and that it is expected that *DQB1* genetic structure at a population level will be influenced by selection. In order to test this hypothesis, the population genetic structure according to *DQB1* locus will be investigated and contrasted to the genetic structure inferred from neutral markers (microsatellites).

II) **To evaluate the functional role of this locus among cetacean populations.** The assessment of *DQB*1 population genetic structure may also indicate ecological factors at which the influence of selection is more prominent. For example differences in the habitat use of sympatric populations or in

geographical regions or parapatric populations may also affect selection pressures. In order to test this hypothesis, the amino acid substitution pattern and nucleotide variation of exon-2 in the *DQB1* locus will be assessed in each population.

5.2 RESULTS

5.2.1 MHC population genetic structure

5.2.1.1 Allele frequencies

The allele and frequencies of the *DQB1* locus for each of the populations of the six focal species were estimated using FSTAT version 2.9.3 (Goudet 2001). The expected allelic frequencies under neutrality were estimated by the Ewens-Watterson-Slatkin exact test. Statistical significance was estimated by Chi-square test ($p<0.05$; Bonferroni correction; see Chapter 2: 2.6.5).

B. acutorostrata and *B. bonaerensis* [Figure 5.2.1 (A,B)]

In the four populations of minke whale, allele frequencies do not deviate significantly from neutral expectations, range of values for chi-square test were x^2=0-1.27, p-value=0.268-1, df=1. There are two alleles present in the Sea of Japan population, M1 and M2.

O. orca [Figure 5.2.3 (A,B)]

Allele frequencies do not deviate significantly from neutral expectations in all putative populations, range of chi-square test values were x^2=0-1.65, p-value=0.199-1, df=1. In Southern Resident (SR), K3 and K4 display the highest frequencies, whilst K2 and K4 are not present.

T. aduncus and *T. truncatus* [Figure 5.2.3 (A,B)]

Allele frequencies are not significantly different from neutrality expectations, range of chi-square test values were x^2=0-1.77, p-value=0.1836-1, df=1. Allele B4 is exclusive in *T. aduncus* while B6 in *T. truncatus*.

M. novaeangliae[Figure 5.2.4 (A,B)]

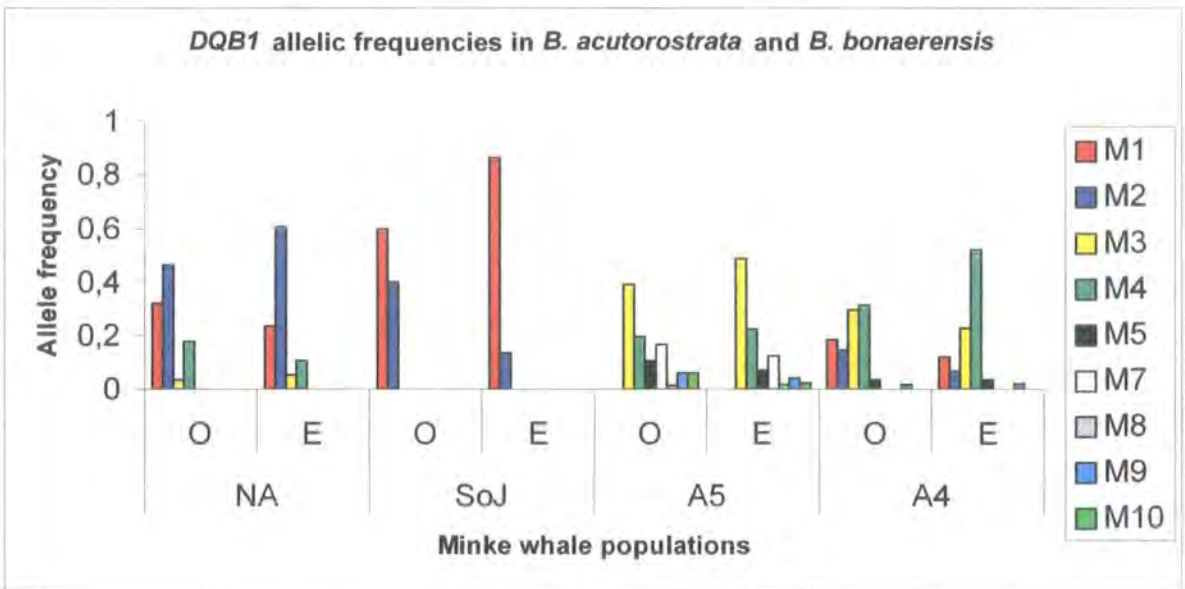
Allelic frequencies do not deviate from neutrality expectations, the range of chi-square test values were x^2=0-0.59, p-value=0.444-1, df=1.

5.2.1.2 Population genetics

In each population of the species, the allelic richness, gene diversity and *Fis* index (Weir and Cockerham 1984) along with the associated probabilities were estimated by FSTAT v. 2.9.3 (Goudet 2001). The Hardy-Weinberg equilibrium indices and statistical significance,

Figure 5.2.1 (A, B): Shows the observed and expected (under neutrality) frequencies of the M1-M10 *DQB1* alleles identified in the present study, in *B. acutorostrata* and *B. bonaerensis* in the North Atlantic (NA), North Pacific (SOJ), Antarctic Area 4 (A4) and Area 5 (A5) populations. Observed frequencies are indicated by O and expected by E.

A)

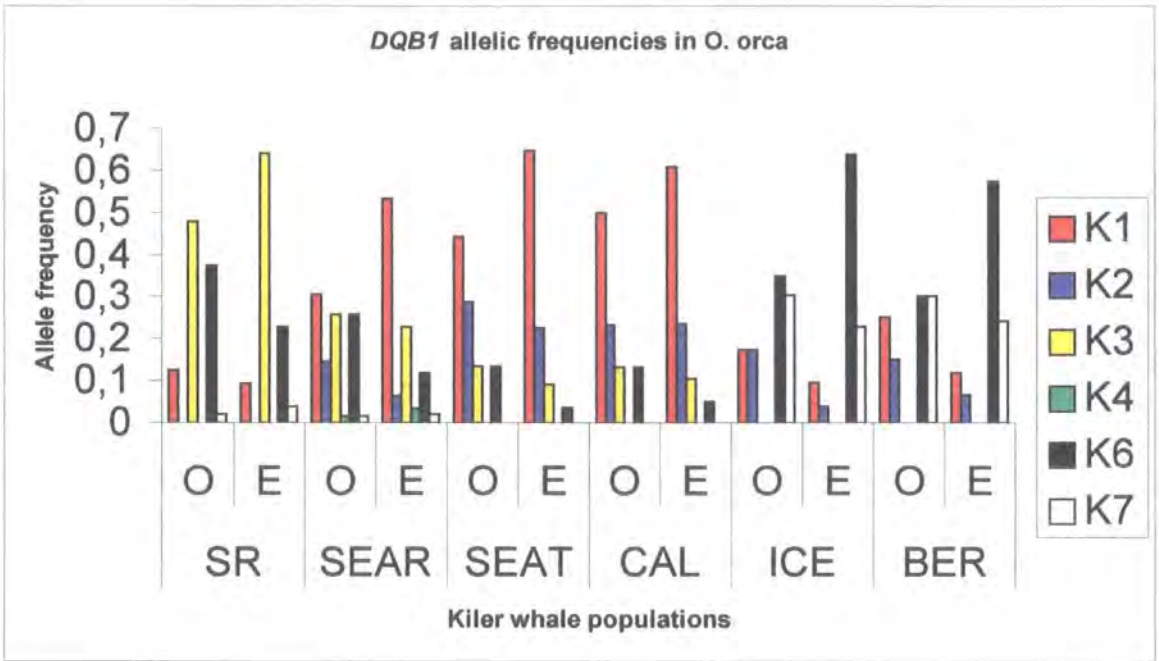


B)

Allele	NA		SOJ		A5		A4	
	O	E	O	E	O	E	O	E
M1	0.321	0.234	0.600	0.864	-	-	0.185	0.121
M2	0.464	0.606	0.400	0.136	-	-	0.148	0.067
M3	0.036	0.052	-	-	0.394	0.491	0.296	0.228
M4	0.179	0.107	-	-	0.197	0.224	0.315	0.523
M5	-	-	-	-	0.106	0.073	0.037	0.037
M7	-	-	-	-	0.167	0.124	-	-
M8	-	-	-	-	0.015	0.018	-	-
M9	-	-	-	-	0.061	0.044	0.019	0.023
M10	-	-	-	-	0.061	0.026	-	-

Figure 5.2.2 (A, B): Shows the observed and expected (under neutrality) frequencies of the K1-K4, K6 and K7 *DQB1* alleles identified in the present study, in *O. orca* in the Southern Residents (SR), South East Alaska Residents (SEAR), South East Alaska Transients (SEAT), California (CAL) of the Eastern North Pacific populations and Iceland residents (ICE) and Berring Sea resident populations (BER). Observed frequencies are indicated by O and expected by E.

A)

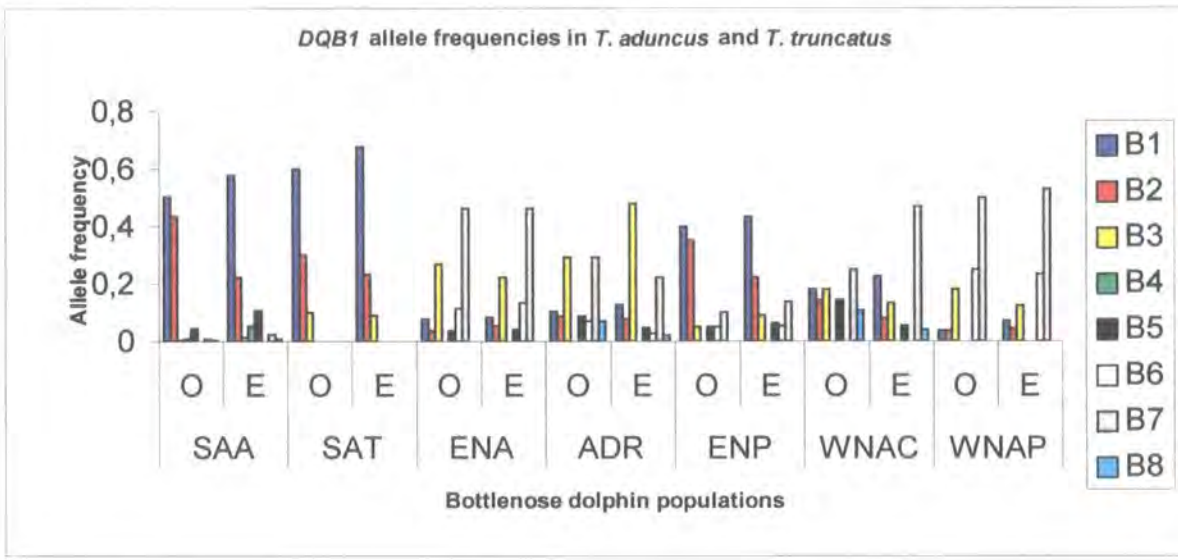


B)

	SR		SEAR		SEAT		CAL		ICE		BER	
	O	E	O	E	O	E	O	E	O	E	O	E
K1	0.125	0.093	0.306	0.534	0.442	0.647	0.500	0.609	0.174	0.095	0.250	0.119
K2	-	-	0.145	0.064	0.288	0.225	0.233	0.235	0.174	0.039	0.150	0.065
K3	0.479	0.641	0.258	0.228	0.135	0.091	0.133	0.106	-	-	-	-
K4	-	-	0.016	0.035	-	-	-	-	-	-	-	-
K6	0.375	0.227	0.258	0.118	0.135	0.036	0.133	0.050	0.348	0.638	0.300	0.575
K7	0.021	0.039	0.016	0.021	-	-	-	-	0.304	0.228	0.300	0.241

Figure 5.2.3 (A, B): Shows the observed and expected (under neutrality) frequencies of the B1-B8 *DQB1* alleles identified in the present study, in *T. aduncus* and *T. truncatus* in the South Africa *T. aduncus* (SAA) and in *T. truncatus,* in South Africa (SAT), Eastern North Atlantic (ENA), Mediterranean (MED), Eastern North Pacific (ENP), Western North Atlantic Coastal (WNAC) and Western North Atlantic Pelagic (WNAP) populations. Observed frequencies are indicated by O and expected by E.

A)

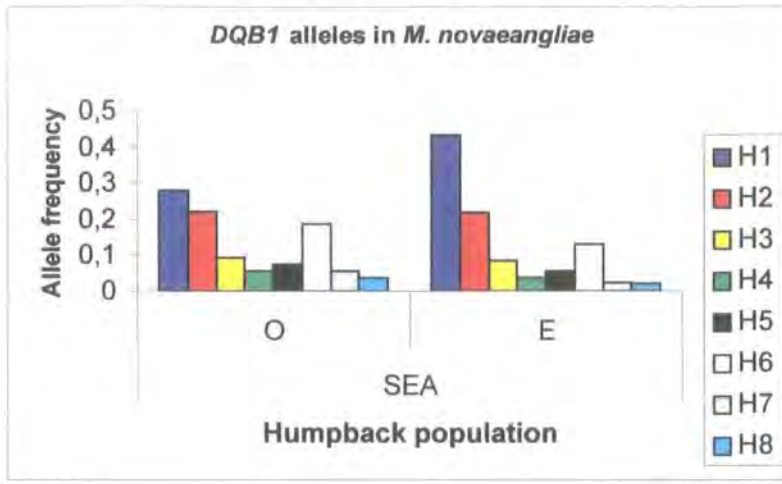


B)

	SAA		SAT		ENA		MED		ENP		WNAC		WNAP	
	O	E	O	E	O	E	O	E	O	E	O	E	O	E
B1	0.504	0.576	0.600	0.677	0.077	0.083	0.103	0.127	0.400	0.433	0.179	0.227	0.036	0.069
B2	0.432	0.223	0.300	0.232	0.038	0.054	0.086	0.076	0.350	0.225	0.143	0.081	0.036	0.043
B3	0.004	0.012	0.100	0.091	0.269	0.225	0.293	0.481	0.050	0.090	0.179	0.132	0.179	0.125
B4	0.007	0.052	-	-	-	-	-	-	-	-	-	-	-	-
B5	0.043	0.106	-	-	0.038	0.041	0.086	0.046	0.050	0.063	0.143	0.052	-	-
B6	-	-	-	-	0.115	0.134	0.069	0.028	0.050	0.052	-	-	0.250	0.235
B7	0.007	0.025	-	-	0.462	0.462	0.293	0.224	0.100	0.138	0.250	0.469	0.500	0.529
B8	0.004	0.006	-	-	-	-	0.069	0.020	-	-	0.107	0.039	-	-

Figure 5.2.4 (A, B): Shows the observed and expected (under neutrality) frequencies of the H1-H8 *DQB1* alleles identified in the present study, in *M. novaeangliae* in South East Alaska samples (SEA). Observed frequencies are indicated by O and expected by E.

A)



B)

	SEA	
	O	E
H1	0.278	0.434
H2	0.222	0.217
H3	0.093	0.084
H4	0.056	0.036
H5	0.074	0.055
H6	0.185	0.130
H7	0.056	0.025
H8	0.037	0.020

and the Ewens-Watterson neutrality test *F* values along with Slatkin's *P value* and *Fst* (Weir and Cockerham 1984) index and Mantel test correlation were estimated by ARLEQUIN version 2.000 (Schneider *et al.* 2000). The expected (under neutrality) and observed values of homozygotes and heterozygotes were estimated by GENEPOP version 3.4 (Raymont and Rousset 1995). In addition, comparison among groups of samples and the estimation of allelic richness, *Ho*, *Hs*, *Fis*, *Fst*, *Relatedness* (Hamilton 1971) and *Corrected Relatedness* was performed by FSTAT v. 2.9.3 (Goudet 2001).

B. acutorostrata and *B. bonaerensis* (Table 5.2.1,i-vi)

Although there is a difference in the numbers of alleles observed between each of the four putative populations of minke whales, one should also consider the difference in the sample sizes. The allelic richness index *R* was estimated in order to correct for sample size differences (see Chapter 2). It is suggested that there is a higher number of alleles present in *B. bonaerensis* than in *B. acutorostrata* and the gene diversity index is higher in A4 and A5 than in NA and SOJ.

An excess of heterozygotes is indicated in all populations by the significance of the P value between observed and expected heterozygosity (Table 5.2.1-i). However, there was no significant difference between the number of expected and observed homozygote and heterozygote individuals (x^2=0.1-2.07, p-value=0.281-0.835, df=1; Table 5.2.1-ii), where the expected values were estimated in accordance to neutrality expectations. In addition, Slatkin's P-value of the F values estimated by the Ewens-Watterson test (Watterson 1978), suggests that none of the four populations are significantly different from neutrality expectations. The *Fis* index value shows that there is a higher degree of outbreeding, and consequently heterozygote excess, in the SOJ population.

The highest *Fst* values result from the pairwise comparison between populations of *B. acutorostrata* and *B. bonaeresis*, an observation consistent with the previous analyses from neutral markers; however, NA and SOJ populations in B. acutorostrata and A4 and A5 in *B. bonaerensis* also exhibit high *Fst* values, *P-values* significantly bigger than zero at $p<0.05$ level (Table 5.2.1-iii). A Mantel test matrix correlation analysis performed between MHC and microsatellite DNA *Fst* values found no significant correlation ($r = 0.429$, $p = 0.235$).

The among-group of samples comparison between the two species of minke whales indicates that allele numbers in *B. bonaerensis* (R=6) is almost twice as big as in *B. acutorostrata* (R=3). Also the relatedness in *B. acutorostrata* is higher than in *B. bonaerensis* an observation which is in conflict with the neutral markers analysis, as mtDNA and microsatellite analysis suggested no genetic divergence among the Antarctic Area VI and V populations (Hoelzel and Dover 1991; Van Pijlen *et al.* 1995).

Table 5.2.1-i: Shows the sample number (n), the observed allele number (A), allelic richness (R), gene diversity (Hs), Hardy-Weinberg expected heterozygosity (He) and observed heterozygosisty (Ho) and the associated p-value, the observed and expected F value by Ewens-Watterson neutrality test and the associated Slatkin's exact p-value, Fis value and associated p-value in the North Atlantic (NA) and North Pacific (SOJ) populations of *B. acutorostrata* and Antarctic Area 4 (A4) and Area 5 (A5) in *B. bonaerensis* populations.

i)

	NA	SOJ	A5	A4
n	14	45	33	27
A	4	2	7	6
R	4	2	6.202	5.290
Hs	0.670	0.483	0.770	0.769
He	0.672	0.621	0.771	0.776
Ho	0.714	0.711	0.818	0.815
P-value	<0.001	0.002	<0.001	<0.001
Fewo	0.352	0.520	0.241	0.245
Fewe	0.474	0.804	0.347	0.381
P-value (slatkin's)	0.285	0.085	0.111	0.186
Fis	-0.066	-0.473	-0.062	-0.060
P-value	0.770	0.998	0.843	0.780

Table 5.2.1-ii: Shows the number of observed (O) and expected under neutrality (E) homozygote and heterozygote individuals in *B. acutorostrata* and *B. bonaerensis*.

ii)

	Homozygotes		Heterozygotes	
	O	E	O	E
North Atlantic	4	4.593	10	9.407
Sea of Japan	13	23.157	32	21.843
Antarctic A5	6	7.554	27	25.446
Antarctic A4	5	6.226	22	20.774

Table 5.2.1-iii: Shows the Fst values of *DQB1* alleles (below diagonal) and microsatellite alleles (above diagonal) in *B. acutorostrata* and *B. bonaerensis* populations.

iii)

	NA	A4	A5	SOJ
NA	-	0.33*	0.33*	0.19*
A4	0.102*	-	nd	0.24*
A5	0.235*	0.056**	-	0.24*
SOJ	0.079**	0.256*	0.381*	-

Figure 5.2.5: Correlation between the extend of genetic differentiation (*Fst*) at microsatellite (MS) and MHC *DQB1* locus (MHC) in minke whale populations (NA, SOJ, A4 and A5). The R-value represents the correlation coefficient and the P-value the statistical significance of Mantel test matrix correlation.

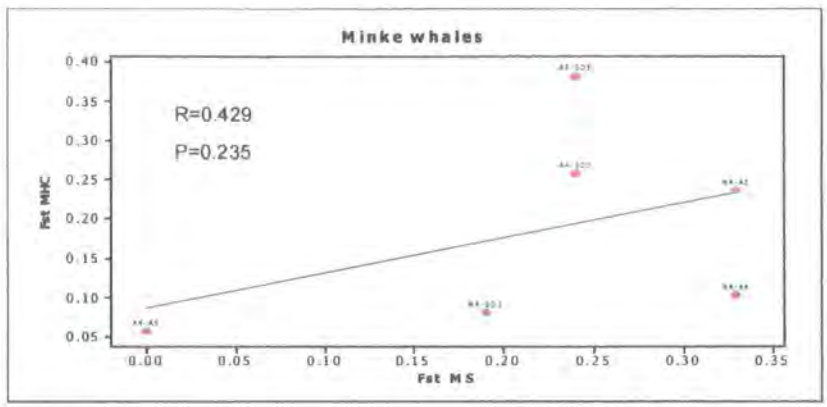


Table 5.2.1-vi: Shows the sample number (n), the observed allele number (A), allelic richness (R), observed heterozygosity (He), gene diversity (Hs), Fis, Fst, Relatedness (Rel) and inbreeding corrected Relatedness (Rel') estimated by the among-group of samples comparison between *B. acutorostrata* and *B. bonaerensis*.

vi)

	B. acutorostrata	***B. bonaerensis***
n	59	60
A	4	9
R	3	5.746
Ho	0.712	0.817
Hs	0.526	0.770
Fis	-0.354	-0.061
Fst	0.086	0.057
Rel	0.226	0.115
Rel'	0.523	0.115

O. orca (Table 5.2.2, i-iv)

The SEAR population has the highest number of alleles with the allelic richness correction (R=5) while all the rest putative populations show a similar number of alleles. In gene diversity (*Hs*), however, the highest values are observed in the BER,

SEAR and ICE putative populations whilst SR, SEAT and CAL also suggest elevated levels of diversity but at a smaller scale.

An excess of heterozygotes is observed in all populations, which significantly deviate from Hardy-Weinberg expectations, as all P-values are <0.001. There was no significant difference between observed and expected under neutrality numbers of homozygotes and heterozygotes (x^2=0.03-3.13, p-value=0.082-0.858, df=1).

However, the SEAT and BER populations significantly deviate from neutrality according to the Ewens-Watterson neutrality test at 0.05 significance level and the ICE populations at the 0.01 significance level, whilst, SR, SEAR and CAL conform to neutrality expectations.

There is no significant difference between the pairwise comparison of populations SEAT-CAL, ICE-BER, SEAR-CAL and SEAT-SEAR. Negative values of *Fst* are due to surrounding errors in the calculations and are interpreted as zero (Long 1986; Jaramilo *et al.* 2001). The population genetic structure in Killer whales follows, to an extend, the population genetic structure obtained by neutral markers studies (see Table 5.2.2-iii). However, Mantel test showed no significant correlation between MHC and MS *Fst* data matrix (r =0.002 p-value =0.486).

Table 5.2.2-i: Shows the sample number (n), the observed allele number (A), allelic richness (R), gene diversity (Hs), Hardy-Weinberg expected heterozygosity (He) and observed heterozygosisty (Ho) and the associated p-value, the observed and expected F value by Ewens-Watterson neutrality test and the associated Slatkin's exact p-value, Fis value and associated p-value in Southeastern Alaska residents (SEAR), Southern residents (SR), Californian transients (CAL), Berring Sea (BER) in North Western Pacific and in Iceland, in the North Eastern Atlantic (ICE) populations.

i)

	SR	SEAR	SEAT	CAL	ICE	BER
n	24	31	26	15	23	10
A	4	6	4	4	4	4
R	3.386	4.622	3.950	3.985	3.988	4.000
Hs	0.624	0.762	0.695	0.676	0.737	0.767
He	0.701	0.789	0.756	0.743	0.742	0.774
Ho	0.750	0.871	0.846	0.867	0.957	0.900
P-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Fewo	0.386	0.248	0.315	0.340	0.274	0.265
Fewe	0.518	0.393	0.523	0.478	0.514	0.440
P-value (slatkin's)	0.358	0.315	0.033**	0.072	0.009*	0.011**
Fis	-0.202	-0.143	-0.217	-0.282	-0.298	-0.174
P-value	0.943	0.953	0.985	0.993	0.998	0.925

Table 5.2.2-ii: Shows the number of observed (O) and expected under neutrality (E) homozygote and heterozygote individuals in *O. orca* populations.

ii)

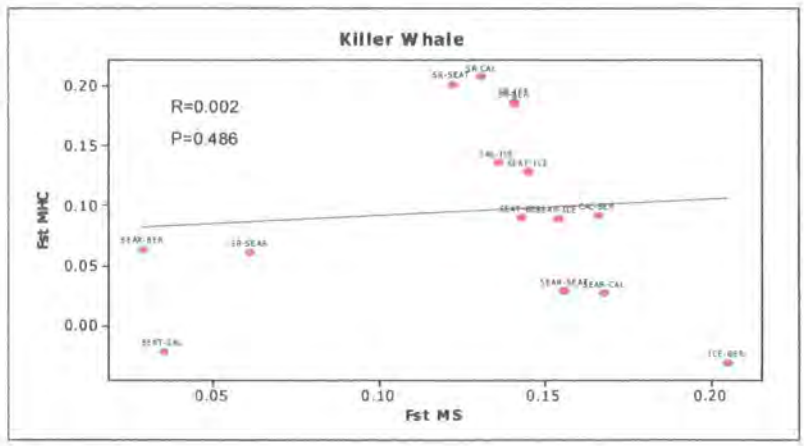
	Homozygotes		Heterozygotes	
	O	E	O	E
SR	6	8.958	18	15.043
SEAR	4	7.328	27	23.672
SEAT	4	7.843	22	18.157
CAL	2	4.759	13	10.241
ICE	1	5.933	22	17.067
BER	1	2.263	9	7.737

Table 5.2.2-iii: Shows the *Fst* values of *DQB1* alleles (below diagonal) and microsatellite alleles (above diagonal) in *O. orca* populations.

iii)

	SR	SEAR	SEAT	CAL	ICE	BER
SR	-	0.061*	0.122*	0.131*	0.141*	0.089*
SEAR	0.061*	-	0.116*	0.168*	0.154*	0.029*
SEAT	0.201*	0.029	-	0.035**	0.145*	0.143*
CAL	0.207*	0.027	-0.022	-	0.136*	0.166*
ICE	0.187*	0.089*	0.128*	0.136*	-	0.205*
BER	0.185*	0.063*	0.090*	0.092**	-0.031	-

Figure 5.2.6: Correlation between the extend of genetic differentiation (*Fst*) at microsatellite (MS) and MHC *DQB1* locus (MHC) in killer whale populations (SR, SEAR, SEAT, CAL, ICE and BER). The R-value represents the correlation coefficient and the P-value the statistical significance of Mantel test matrix correlation.



For the purpose of among group comparison, killer whale samples were grouped into South East Alaskan residents (SR+SEAR) and transients (SEAT+CAL). All groups display high values of heterozygosity and gene diversity. *Fst* values are relatively small where the highest value (0.065) is showed by SR+SEAR. There is a significant difference between the Relatedness and Corrected relatedness of SEAR+CAL and thus, the corrected relatedness value are considered.

Table 5.2.2v: Shows the sample number (n), the observed allele number (A), allelic richness (R), observed heterozygosity (He), gene diversity (Hs), Fis, Fst, Relatedness (Rel) and inbreeding corrected Relatedness (Rel') estimated by the among-group of samples comparison between the Eastern North Pacific residents (SR and SEAR) and transients (SEAT and CAL) and North Western Pacific and North Eastern Atlantic resident (ICE and BER) populations in *O. orca*.

v)

	SR+SEAR	SEAT+CAL
n	55	41
A	6	4
R	4.004	3.967
Ho	0.818	0.854
Hs	0.702	0.689
Fis	-0.166	-0.240
Fst	0.064	-0.016
Rel	0.140	-0.042
Rel'	0.284	0.387

T. aduncus and *T. truncatus* (Table 5.2.3,i-iv)

The *T. aduncus* and Mediterranean *T. truncatus* populations display the highest number of alleles; however, allelic richness indicates that there is an excess in allele number in *T. aduncus*. ENA and ENP show the same number whilst the SAT population had the lowest value. WNAC shows a slightly bigger number of alleles compared to WNAP. All populations exhibit a relatively large value of gene diversity with MED and WNAC having the highest values and SAA and SAT the lowest values.

Table 5.2.3-i: Shows the sample number (n), the observed allele number (A), allelic richness (R), gene diversity (Hs), Hardy-Weinberg expected heterozygosity (He) and observed heterozygosisty (Ho) and the associated p-value, the observed and expected F value by Ewens-Watterson neutrality test and the associated Slatkin's exact p-value, Fis value and associated p-value in *T. truncatus* populations from MEDaitic (MED), South Africa (SA), Eastern North Atlantic (ENA), Eastern North Pacific (ENP), Western North Atlantic Pelagic (WNAP) and Western North Atlantic Coastal (WNAC) and *T. aduncus* South Africa population.

i)

	SAA	SAT	ENA	MED	ENP	WNAC	WNAP
n	140	10	13	29	10	14	14
A	7	3	6	7	6	6	5
R	3.015	3.000	5.485	6.364	6.000	5.975	4.428
Hs	0.559	0.556	0.718	0.807	0.728	0.857	0.673
He	0.609	0.568	0.800	0.828	0.737	0.899	0.741
Ho	0.657	0.800	0.769	0.834	0.800	0.714	0.786
Pvalue	0.012	0.367	0.076	<0.001	0.011	<0.001	0.059
Fewo	0.442	0.460	0.308	0.207	0.300	0.179	0.347
Fewe	0.440	0.559	0.319	0.336	0.293	0.326	0.389
Pvalue (slatkin's)	0.887	0.291	0.564	0.012	0.749	0.001	0.540
Fis	-0.175	-0.440	-0.071	-0.026	-0.237	0.167	-0.167
Pvalue	0.997	1.000	0.777	0.708	0.950	0.149	0.893

Not all populations show a heterozygote excess, as populations SAT, ENA and WNAP are not significantly different from Hardy-Weinberg expectations. WNAC population shows a heterozygote deficiency. The observed and expected homozygote and heterozygote individual numbers further illustrate this, as observed homozygotes are slightly fewer than expected and observed heterozygotes are slightly more than expected, with the exception of WNAC. No significant difference between observed and expected under neutrality numbers of homozygotes and heterozygotes (x^2=0.08-1.28, p-value=0.258-0.782, df=1). The highest negative *Fis* value is showed by the SAA population while SAA, ENA, MED, ENP and WNAP also suggest an excess in heterozygosity. The only positive value is by WNAC, which suggests inbreeding. In addition, the WNAC population is also the only population, which is significantly different from neutrality expectations.

Fst values between the population pairwise comparisons of SAA-SAT-ENP, ENA-MED-WNAC and ENP-WNAC are not significantly different; on the contrary, the MS data suggest that *Fst* values for these populations are significantly different

(Natoli *et al.* 2004). There is no available *Fst* data for the pairwise comparisons of SAA-SAT. There was no significant correlation between the MHC and MS data matrix (r =0.059, p =0.398).

Table 5.2.3-ii: Shows the number of observed (O) and expected under neutrality (E) homozygote and heterozygote individuals in *T. aduncus* and *T. truncatus* populations.

ii)

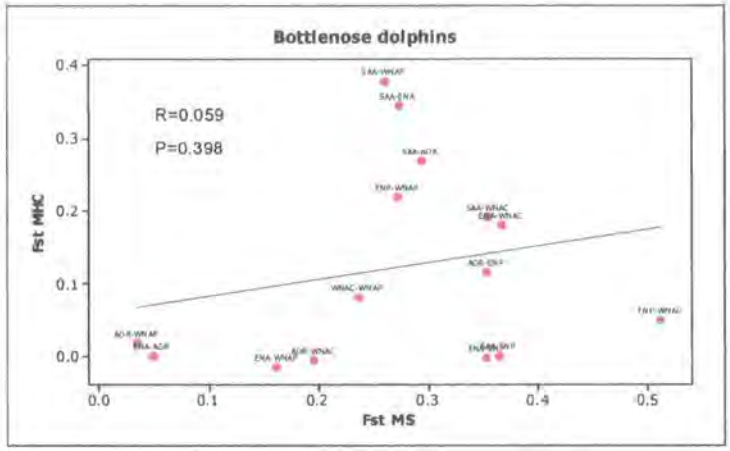
	Homozygotes		Heterozygotes	
	O	E	O	E
SAA	48	61.642	92	78.358
SAT	2	4.316	8	5.684
ENA	3	3.640	10	9.360
MED	5	5.597	24	23.404
ENP	1	2.632	9	7.368
WNAC	4	2.074	10	11.926
WNAP	3	4.518	11	9.481

Table 5.2.3-iii: Shows the Fst values of *DQB1* alleles (below diagonal) and microsatellite alleles (above diagonal) in *T. aduncus* and *T. truncatus* populations.

iii)

	SAA	SAT	ENA	MED	ENP	WNAC	WNAP
SAA	-	nd	0.273*	0.293*	0.364*	0.354*	0.260*
SAT	0.007	-	nd	nd	nd	nd	Nd
ENA	0.345*	0.292*	-	0.048**	0.353*	0.367*	0.161**
MED	0.269*	0.203*	<0.001	-	0.353*	0.196*	0.034*
ENP	0.001	-0.004	-0.002*	0.116*	-	0.511*	0.272*
WNAC	0.191*	0.140**	0.181	-0.006	0.050	-	0.236*
WNAP	0.378*	0.340*	-0.015	0.018**	0.218*	0.081**	-

Figure 5.2.7: Shows the correlation between the extend of genetic differentiation (*Fst*) at microsatellite (MS) and MHC *DQB1* locus (MHC) in bottlenose dolphins populations (SAA, SAT, ENA, MED, ENP, WNAC and WNAP). The R-value represents the correlation coefficient and the P-value the statistical significance of Mantel test matrix correlation.



M. novaeangliae (Table 5.2.3,i-v)

A homozygote excess is observed as *Ho* is smaller than *He* and it deviated significantly from Hardy-Weinberg equilibrium. However, one should consider the fact that samples were collected in South East Alaska where Humpback whales form temporary feeding associations and thus, the Wahlund effect is a possible explanation for the Hardy-Weinberg deviation observed (Gillespie 1998). In addition, the positive value of *Fis*, may indicate inbreeding. The numbers of observed and expected under neutrality homozygote and heterozygote individuals are not significantly different (x^2=0.05-0.08, p-value=0.783-0.823, df=1). However, neutrality is rejected, as Slatkin's P-value is 0.024.

Table 5.2.4-i: Shows the sample number (n), the observed allele number (A), allelic richness (R), gene diversity (Hs), Hardy-Weinberg expected heterozygosity (He) and observed heterozygosisty (Ho) and the associated p-value, the observed and expected F value by Ewens-Watterson neutrality test and the associated Slatkin's exact p-value, Fis value and associated p-value in *M. novaeangliae* southeast Alaska (SEA) samples.

i)

	SEA
n	27
A	8
R	8
Hs	0.834
He	0.837
Ho	0.778
Pvalue	0.027
Fewo	0.182
Fewe	0.290
Pvalue (slatkin's)	0.024
Fis	0.067
Pvalue	0.310

Table 5.2.4-ii: Shows the number of observed (O) and expected under neutrality (E) homozygote and heterozygote individuals in *M. novaeangliae* southeast Alaska (SEA) samples.

ii)

	Homozygotes		Heterozygotes	
	O	**E**	**O**	**E**
SEA	6	4.509	21	22.491

5.2.2 Population-specific amino acid substitution patterns and structural analysis of *DQB1*-PBR

Association between population-specific polymorphism and structural patterns of the PBR region in class II loci and susceptibility or resistance to infectious diseases has been suggested in several studies on vertebrates (*e.g.* Verspoor and Jordan 1989; Hill *et al.* 1991; Boyce *et al.* 1997; Sanjeevi *et al.* 1999; Landry and Bernatchez 2001; Cohen 2002). It has been shown, in humans, bovines and fish species that differences in selection pressure experienced by different populations may also result in differences in the amino acid substitution pattern in class II loci (Hill *et al.* 1991; Boyce *et al.* 1997;

Landry and Bernatchez 2001; Cohen 2002). Furthermore, it has been indicated, through site-directed mutagenesis in *DQ* and *DR* alleles of HLA, that selective peptide binding is greatly affected by the charges of Pocket 4 amino acid residues (β70, β71 and β74) in the PBR, as when the total charge of the pocket changes then different peptides are presented by the MHC molecule (Ou *et al.* 1992; Wecherpfennig and Strominger 1995; Hammer *et al.* 1995). This has resulted in the categorization of MHC alleles into four charge types: positive (+), di-charged (+/-), negative (-) and neutral (n; see Chapter 2 for charge estimation; Ou *et al.* 1998). The purpose of the structure analysis is to evaluate: a) the amino acid substitution patterns among different populations and b) differences in the allele charge character among populations.

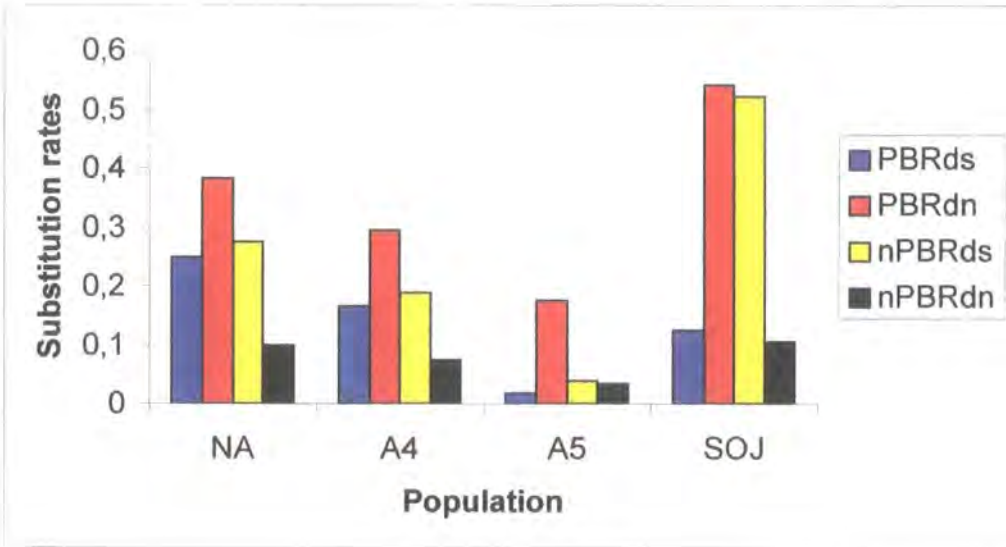
5.2.2.1 Population substitution patterns

The Nei-Gojobori method (MEGA version 2.1; Kumar *et al.* 2001) was used in order to estimate the rates of non-synonymous substitutions in non-synonymous sites (*dn*) and synonymous substitutions in synonymous sites (*ds*) in the Peptide Binding Region (PBR) and the non-Peptide Binding Region (nPBR).

Populations of *B. bonaerensis*, *B. acutorostrata*, *O. orca*, *T. aduncus*, *T. truncatus* and *M. novaeangliae* (Figures 5.2.8-11 A, B), show ω ratios at the PBR significantly greater than one. There was no significant difference between any of the populations in *O. orca* (range of t-values=0-0.62, p-values=0.538-1, df=26-PBR and 114 for nPBR). In minke whale populations, Antarctic area 5 population (*B. bonaerensis*) showed a significantly higher ω ratio in the PBR sites from the North Atlantic (t-value=8.32, p-value=<0.001, df=26; Bonferroni correction α=0.0125) and Sea of Japan population (t-value=7.96, p-value=<0.001, df=26, Bonferroni correction α=0.0125; *B. acutorostrata*).

Figure 5.2.8 (A, B): Shows the rate of synonymous substitutions per synonymous sites (*ds*) and non-synonymous substitutions per non-synonymous sites (*dn*) in the putative Peptide Binding Region (PBR) and the non Peptide Binding Region (nPBR) of the *DQB1* alleles in each of the *B. acutorostrata* and *B. bonaerensis* populations.

A)

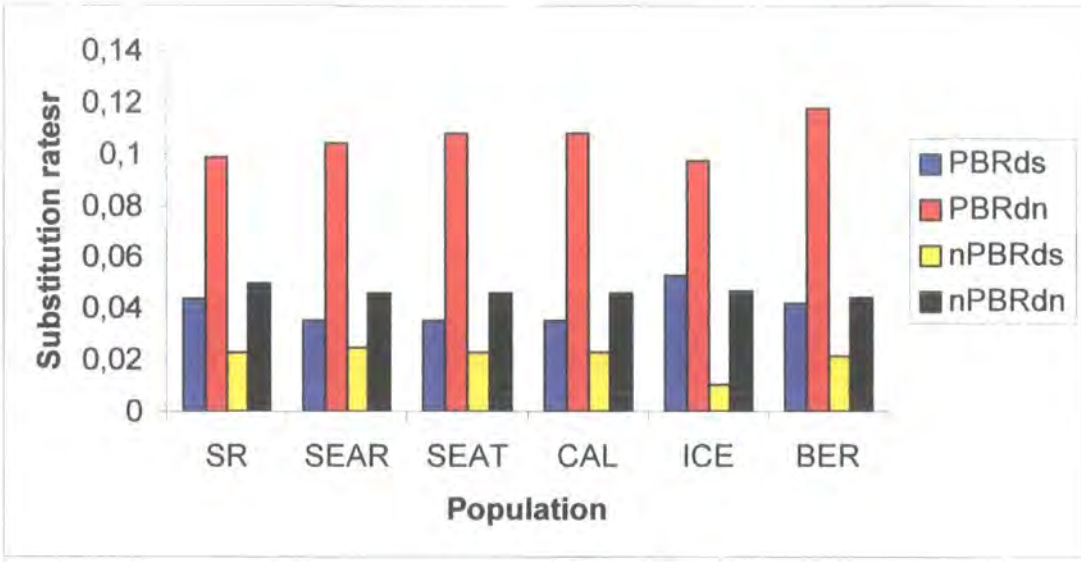


B)

	NA	A4	A5	SOJ
FULLds	0.059	0.045	0.031	0.024
	(0.029	(0.023)	(0.018)	(0.025)
FULLdn	0.135	0.108	0.070	0.156
	(0.035)	(0.028)	(0.021)	(0.047)
ω	**2.288**	**2.400**	**2.258**	**6.500**
PBRds	0.2489	0.166	0.0176	0.1250
	(0.0711)	(0.0464)	(0.0112)	(0.1388)
PBRdn	0.3824	0.2944	0.1763	0.5417
	(0.1304)	(0.0916)	(0.0833)	(0.0843)
ω	**1.536**	**1.773**	**10.017**	**4.333**
nPBRds	0.2747	0.1833	0.0385	0.5214
	(0.0847)	(0.0407)	(0.0190)	(0.1787)
nPBRdn	0.0987	0.0749	0.0334	0.1059
	(0.0263)	(0.0239)	(0.0199)	(0.0462)
ω	**0.359**	**0.407**	**0.867**	**0.203**

Figure 5.2.9 (A, B): Shows the rate of synonymous substitutions per synonymous sites (*ds*) and non-synonymous substitutions per non-synonymous sites (*dn*) in the putative Peptide Binding Region (PBR) and the non Peptide Binding (nPBR) region of the *DQB1* alleles in each of the *O. orca* populations.

A)

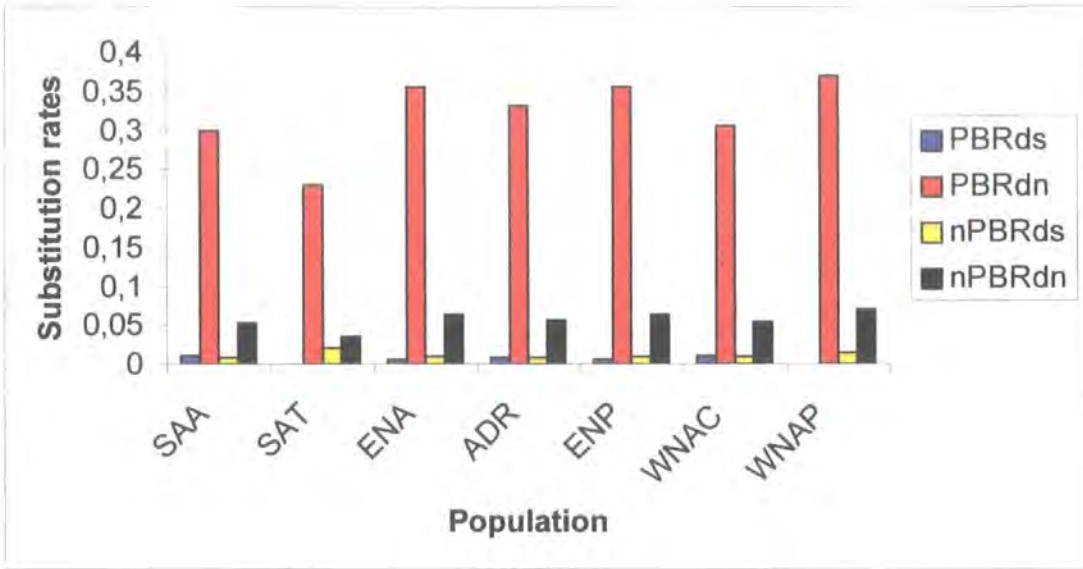


B)

	SR	SEAR	SEAT	CAL	ICE	BER
FULLds	0.028	0.027	0.025	0.025	0.020	0.026
	(0.018)	(0.017)	(0.016)	(0.016)	(0.014)	(0.016)
FULLdn	0.062	0.060	0.061	0.061	0.059	0.062
	(0.021)	(0.020)	(0.020)	(0.020)	(0.022)	(0.021)
ω	**2.214**	**2.222**	**2.440**	**2.440**	**2.950**	**2.385**
PBRds	0.0440	0.0352	0.0352	0.0352	0.0529	0.0420
	(0.0517)	(0.0361)	(0.0383)	(0.0369)	(0.0577)	(0.0440)
PBRdn	0.0989	0.1043	0.1081	0.1081	0.0973	0.1177
	(0.0613)	(0.0633)	(0.0589)	(0.0589)	(0.0666)	(0.0569)
ω	**2.248**	**2.963**	**3.071**	**3.071**	**1.839**	**2.802**
nPBRds	0.0232	0.0247	0.0229	0.0229	0.0103	0.0217
	(0.0188)	(0.0198)	(0.0187)	(0.0180)	(0.0108)	(0.0164)
nPBRdn	0.0496	0.0462	0.0459	0.0459	0.0468	0.0442
	(0.0220)	(0.0196)	(0.0207)	(0.0199)	(0.0217)	(0.0196)
ω	**2.138**	**1.871**	**2.004**	**2.004**	**4.544**	**2.037**

Figure 5.2.10 (A, B): Shows the rate of synonymous substitutions per synonymous sites (*ds*) and non-synonymous substitutions per non-synonymous sites (*dn*) in the putative Peptide Binding Region (PBR) and the non Peptide Binding (nPBR) region of the *DQB* alleles in the *T. aduncus* and each of the *T. truncatus* populations.

A)

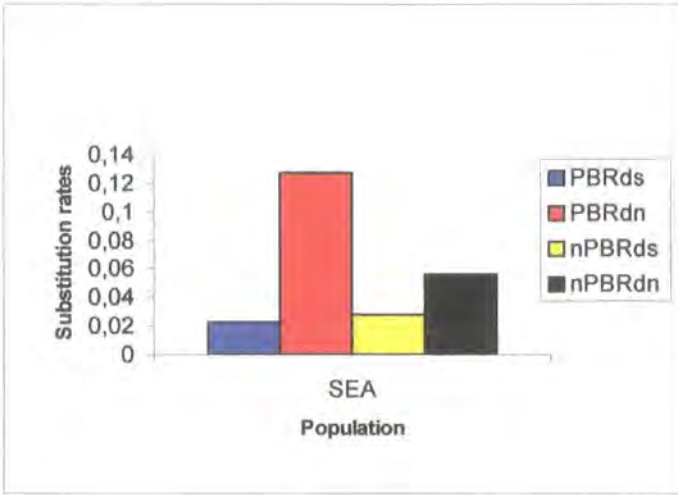


B)

	SAA	SAT	ENA	ADR	ENP	WNAC	WNAP
FULLds	0.0079	0.016	0.009	0.008	0.009	0.010	0.012
	(0.0069)	(0.017)	(0.008)	(0.007)	(0.008)	(0.008)	(0.012)
FULLdn	0.1050	0.077	0.126	0.115	0.125	0.108	0.133
	(0.0243)	(0.024)	(0.028)	(0.026)	(0.027)	(0.025)	(0.028)
ω	**13.292**	**4.813**	**14.000**	**14.375**	**13.889**	**10.800**	**11.083**
PBRds	0.0107	0	0.0057	0.0089	0.0057	0.0114	0
	(0.0085)		(0.0059)	(0.0071)	(0.0059)	(0.0107)	
PBRdn	0.2991	0.2298	0.3561	0.3309	0.3561	0.3051	0.3692
	(0.0876)	(0.0918)	(0.0802)	(0.0755)	(0.0797)	(0.0836)	(0.0767)
ω	**27.953**	**NA**	**62.474**	**37.180**	**62.474**	**26.763**	**NA**
nPBRds	0.0086	0.0203	0.0101	0.0086	0.0101	0.0100	0.0151
	(0.0091)	(0.0209)	(0.0110)	(0.0087)	(0.0111)	(0.0106)	(0.0171)
nPBRdn	0.0530	0.0360	0.0633	0.0561	0.0633	0.0537	0.0701
	(0.0201)	(0.0183)	(0.0204)	(0.0187)	(0.0204)	(0.0184)	(0.0267)
ω	**6.163**	**1.773**	**6.267**	**6.523**	**6.267**	**5.370**	**4.642**

Figure 5.2.11 (A, B): Shows the rate of synonymous substitutions per synonymous sites (*ds*) and non-synonymous substitutions per non-synonymous sites (*dn*) in the putative Peptide Binding Region (PBR) and the non Peptide Binding (nPBR) region of the *DQB1* alleles in M. novaeangliae South East Alaska samples.

A)



B)

	SEA
FULLds	0.0536
	(0.0183)
FULLdn	0.0596
	(0.0274)
ω	**1.112**
PBRds	0.023
	(0.0.23)
PBRdn	0.127
	(0.067)
ω	**5.522**
nPBRds	0.028
	(0.020)
nPBRdn	0.056
	(0.018)
ω	**2.000**

5.2.2.2 Allele charges among populations

Alleles in each population of the focal species were categorized into the four functional categories previously defined in class II *DRB1* and *DQB1* HLA (Ou *et al.* 1998). In Pocket 4 of the PBR of HLA *DRB1* locus (where the most extensive studies have taken place), 50% of the alleles are classified as di-charged (+/-), 39% as positive (+), 13% as negative (-) and 8% as Neutral (n) (Ou *et al.* 1998). In the six cetacean species, 36% of the alleles identified in this study are (+/-), 6% are (+), 53% are (-) and 6% are (n).

Figure 5.2.12 shows the functional distribution of the alleles present in the *B. acutorostrata* and *B. bonaerensis* populations. Allele M10, which is exclusive to the A5 population of *B. bonaerensis* is classified in the (+) group and it is the only positive allele present. In addition, all populations show a higher number of (-) alleles than (+), however, they are not significantly different. No neutral alleles are present.

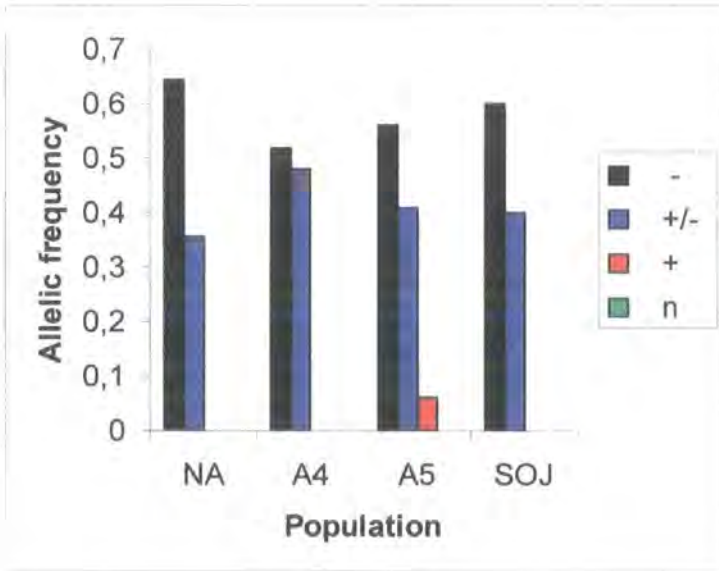
Figure 5.2.13 shows the functional distribution of the alleles present in the *O. orca* populations. There are no neutral alleles present. In populations ICE and BER there were no (+/-) alleles present and (-) alleles were the highest number observed.

Figure 5.2.14 shows the functional distribution of the alleles present in *T. aduncus* and *T. truncatus* populations. There are no (+) alleles present in the SAA and SAT populations. In addition, no (-) alleles are present in SAT. No neutral alleles are present.

Figure 5.2.15 shows the functional distribution of the alleles present in the *M. novaeangliae* population. There are no (+) alleles present; neutral alleles in this study were observed only in this focal species. In fact, the allelic proportions were 70% negative, 6% di-charged and 24% neutral.

Figure 5.2.12 (A, B): Shows the categorization of the *DQB* alleles M1-M10 identified in this study based on the total charge of Pocket 4 amino acid residues (β70, β71 and β74) in each of the population in *B. acutorostrata* and *B. bonaerensis*.

A)

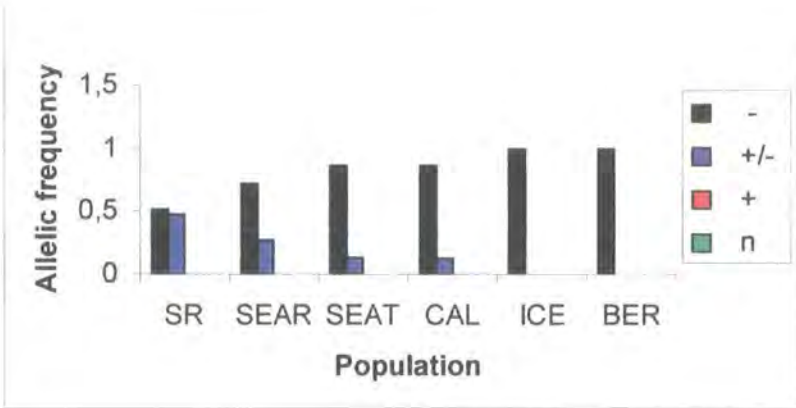


B)

+	+/-	-	n
-	M1	-	-
-	-	M2	-
-	M3	-	-
-	-	M4	-
-	-	M5	-
-	-	M6	-
-	-	M7	-
-	M8	-	-
-	-	M9	-
M10	-	-	-

Figure 5.2.13 (A, B): Shows the categorization of the *DQB* alleles M1-M10 identified in this study based on the total charge of Pocket 4 amino acid residues (β70, β71 and β74) in each of the population in *O. orca*.

A)

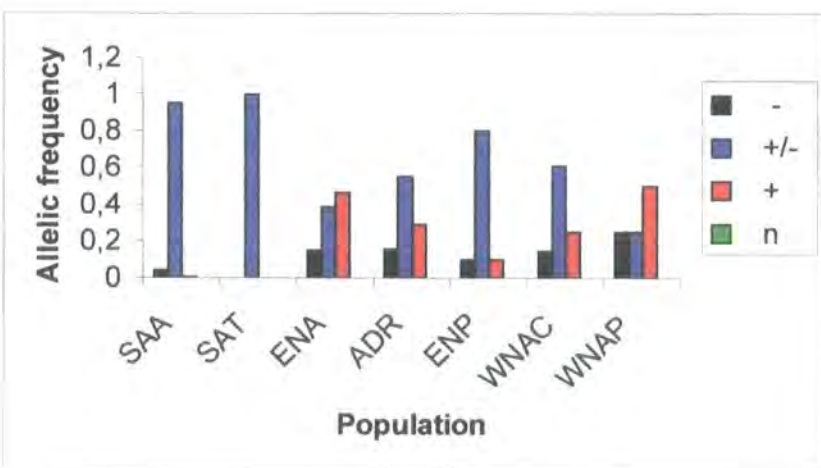


B)

+	+/-	-	n
-	-	K1	-
-	-	K2	-
-	K3	-	-
-	K4	-	-
-	K5	-	-
-	-	K6	-
-	-	K7	-

Figure 5.2.14 (A, B): Shows the categorization of the *DQB* alleles M1-M10 identified in this study based on the total charge of Pocket 4 amino acid residues (β70, β71 and β74) in each of the population in *T. aduncus* and *T. truncates*.

A)

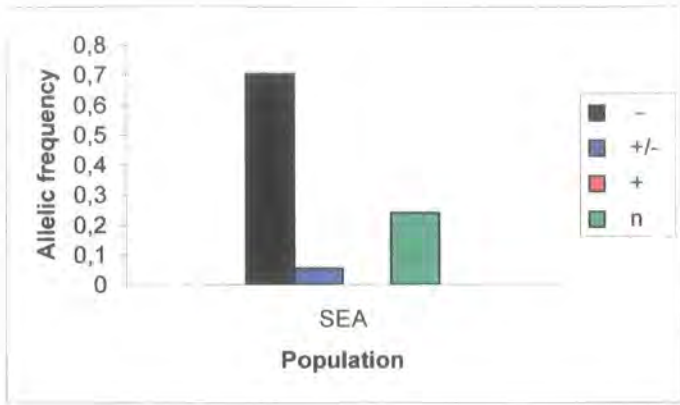


B)

+	+/-	-	n
-	B1	-	-
-	B2	-	-
-	B3	-	-
-	B4	-	-
-	-	B5	-
-	-	B6	-
B7	-	-	-
-	B8	-	-

Figure 5.2.15 (A, B): Shows the categorization of the *DQB* alleles M1-M10 identified in this study based on the total charge of Pocket 4 amino acid residues (β70, β71 and β74) in *M. novaeangliae*.

A)



B)

+	+/-	-	n
-		H1-	-
-		H2	-
-		H3	-
-	H4	-	-
-	-	H5	-
-	-		H6-
	-	-	H7
-		H8-	-

5.3 DISCUSSION

Chapters 3 and 4 showed high rates of non-synonymous substitutions in the Peptide Binding Region (PBR) for each of the studied species, which indicates that selection has increased diversity at sites that are involved in antigen presentation. This is confirmed at the population level, as all populations show a significantly higher rate of non-synonymous substitutions to synonymous in the PBR compared to the non-PBR. In addition, 89% of the populations sampled had significantly higher heterozygocity than expected under Hardy-Weinberg equilibrium, an observation consistent with the expectations of balancing selection and/ or heterosis. It has been shown by Penn *et al.* (2002) and Kurtz *et al.* (2004) that heterozygotes may have a selective advantage when under multiple pathogen infection and heterozygote excess in MHC has been attributed to heterozygote superiority. However, although in this study homozygote deficiency was observed in the majority of the population, the homozygosity levels were not significantly lower than the levels inferred by neutrality, with the exception of the Mediterranean and Western North Atlantic Coastal population in *T. truncatus*, *M. novaeangliae* and the South Eastern Alaska Transient, Icelandic and the Berring Sea

populations in *O. orca*. In these populations, neutrality was rejected by the Ewens-Watterson neutrality test. This result has also been observed in human populations where in a study performed by Valdes *et al.* (1999) in HLA *DQB1*, neutrality was rejected in 9 populations out of 20, and a similar pattern has been observed in class II loci. This observation has been attributed to directional selection of single favored alleles rather than balancing selection (Sanchez-Mazas *et al.* 2000; Meyer and Thomson 2001). In addition, the Ewens-Watterson test is suggested to be not very powerful when the number of alleles in the sample is relatively small, thus failure to reject neutrality cannot rule out that selection is taking place (Fernandez-Vinã *et al.* 1997; Meyer and Thomson 2001).

MHC selection, and thus evolution, can take place, both, at very long and relatively short time scales, depending on population genetic parameters such as migration and isolation (*e.g.* Klein 1986; Wills 1991; Garrigan and Hedrick 2003). For example, the trans-species pattern (Klein 1989) is the result of long time-scale evolution (Klein *et al.* 1998). Local variation in alleles, such as those which confer protection to severe malaria in African populations, are the result of short time scale evolution (Hill *et al.* 1991; De-Campos Lima *et al.* 1993). Therefore, it has been suggested that allelic variation and frequency differences may reflect the pattern of antigenic stressors in local environments (Meyer and Thomson 2001). According to the *Fst* analysis performed in this study, there is population differentiation inferred by *DQB1* locus population genetics, an observation which may indicate differences in the antigenic stressors experienced by each population. In addition, the data matrix inferred by the *Fst* analysis in *B. acutorostrata*, *B. bonaerensis*, *T. aduncus*, *T. truncatus* and *O. orca* is significantly different from the data matrix inferred by microsatellite DNA analyses, consistent with expectations that population differentiation at the *DQB1* locus would be shaped by selection as well as genetic drift and migration. However, it should be noted that small sample sizes greatly affect allelic frequencies and *Fst* values accordingly (Long 1986; Jaramillo *et al.* 2001).

B. acutorostrata and *B. bonaerensis*

Studies on neutral markers indicate that samples taken from Antarctic management Area 4 (A4) and Antarctic management Area 5 (A5) putative populations where not significantly differentiated (VanPijlen *et al.* 1995). However, my study using *DQB1* locus does show a significant difference ($p<0.05$) between A4 and A5

populations. This may be attributed to the fact that the microsatellite study was based on relatively small number of microsatellite loci (6) and sample size (24 and 21 in A4 and A5 respectively) and therefore there was low power. In addition, charge analysis showed that allele M10 which is exclusive in the A5 population is also the only positively charged allele present in the samples included in the present study. However, the main observation of the MHC population genetics analysis is the low number of alleles (R=3; Table 5.2.1-iv) present in *B. acutorostrata*.

Whaling operations still take place on both species of minke whales and although population stocks are considered to be in a far better condition than the rest of baleen whales, there are concerns regarding the population status and the effects of continued whaling. In *B. acutorostrata*, in particular, the Sea of Japan population was protected under IWC regulations in 1987 due to considerable depletion. In late 1994, Japanese whaling began again in the North Pacific with a nationally established limit of 100 animals annually. Studies in mammalian species which have gone through a significant genetic variation reduction (bottleneck effect) show small numbers of *DQB1* alleles (Ellegren *et al.* 1993; Mikko and Andersson 1995; Hoelzel *et al.* 1999; Van Der Valt *et al.* 2001). The small number of *DQB1* alleles present may suggest that the Sea of Japan population is significantly depleted. However, further research is required in order for any conclusions to be drawn.

O. orca

The sympatric populations of killer whales in the North Eastern Pacific have been categorized according to their dietary and seasonal movement to either Residents or Transients (see 5.1.2.1). Residents show strong seasonal movement according to the coastal migration of salmon, as they feed primarily on salmon species and also on a variety of non-salmonid fish species (Ford *et al.* 1998; Ford 2002); nevertheless, interaction does occur as residents have been reported to kill porpoises and seals on rare occasions. Transients show a very restricted seasonal movement and feed on marine mammal species, such as harbor seals, Steller sea lions, California sea lions, Dall's porpoises, Pacific white-sided dolphins (Ford and Ellis 1999).

The MHC *Fst* values follow to an extend the differentiation pattern observed in microsatellite analysis where the Southern Residents (SR), South East Alaskan Residents (SEAR), South East Alaskan Transients (SEAT) and California (CAL) North Eastern Pacific populations are significantly different from the North Eastern Atlantic

(ICE) and North Western Pacific (BER) populations. However, according to the MHC analysis, the *Fst* values of SEAR and SEAT (0.029) SEAR and CAL (0.027) SEAT and CAL (-0.022) and ICE and BER (-0.031) are not significantly different.

According to microsatellite DNA, residents and transients in North Easten Pacific and the resident populations in the North Eastern Atlantic and North Western Pacific are genetically distinct. However, this in not supported by the MHC analysis. One of the main characteristics of MHC evolution is the persistence of alleles and allelic frequencies through evolutionary time and through balancing selection and heterosis (e.g. Klein 1986; Wills 1992; Garrigan and Hedrick 2003). Therefore, in transient and resident populations, alleles may have persisted predating the divergence between these two sympatric populations. Selection on *DQB1* alleles may have also maintained the allelic frequencies observed and this may indicate that resident and transients are exposed to similar antigenic stressors and due to this, MHC *Fst* values show no differentiation in *DQB1* alleles. However, the Southern resident population (SR) of North Eastern Pacific is significantly different from SEAR, SEAT and CAL an observation which is inconsistent to the expectations of the above hypothetical interpretation.

The Southern Resident population size was estimated to 89 individuals in 1998 (Braid 2001) and is one of the most studied populations in killer whales. The last three years the population size has declined by 10% due to an increased level of mortalities, particularly of adult females. A constantly decreasing population size can have an affect on the estimation of the *Fst* value (Gillspie 1998), therefore, this may suggest a possible reason for the population differentiation of SR to SEAR, SEAT and CAL.

T. aduncus and T. truncatus

The population structure inferred by MHC in bottlenose dolphins may also be attributed to *DQB1* allelic persistence through evolutionary time due to selection. There is no significant difference between the South Africa populations of *T. aduncus* (SAA) and *T. truncatus* (SAT) and the Eastern North Pacific *T. truncatus* population (ENP), although the small sample size of SAT (n=10) should be noted. Further to *T. truncatus* populations, the Eastern North Atlantic (ENA) population is not significantly different to the Mediterranean (MED), Western North Atlantic Coastal (WNAC) and Pelagic (WNAP) populations. However, MED is significantly different to the WNAP and not to WNAC. Eastern North Pacific (ENP) population is not significantly different to

WNAC. In addition, WANC and WNAP populations are significantly different (0.081) according to microsatellite data expectations.

In the Western Atlantic there are two parapatric *T. truncatus* morphotypes classified as coastal and pelagic, as comparisons between these two populations at both nuclear and mitochondrial DNA markers indicated a distinct genetic differentiation (Hoelzel *et al.* 1998). Parasite load, feeding habits and body size also differentiate between these two populations, as the pelagic morphotype is relatively smaller, feeds on coastal fish species and presents a significant higher parasite load, whilst the pelagic form is larger and feeds on pelagic fish species (Mead and Potter 1995; Hoelzel *et al.* 1998). The level of genetic variation observed in the coastal form is significantly less than the level observed in the pelagic, an observation which has been attributed to a founder event originating from the larger pelagic population (Hoelzel *et al.* 1998). However, according to *DQB1* locus, there is a higher number of alleles present in the coastal population than in the pelagic population, an observation which may be attributed to differences in parasite load. In addition, coastal environments are rich in human pathogenic organisms (see review by Griffin *et al.* 2003) and hypothetically the coastal population is under higher pathogenic pressures than the pelagic population, although further research is required in order for any conclusions to be drawn.

5.3.1 Functional categorization of population-specific *DQB1* alleles

Recent elucidation of the molecular structure of MHC class II loci has identified subregions (pockets) within the Peptide Binding Region (PBR), which greatly affect the binding and therefore the presentation and recognition of a given antigen by T-cells. These pockets may determine to a great extend which antigens will be recognized and which will not (Stern *et al.* 1994; Fu *et al.* 1995; Ou *et al.* 1996). There are six subregions in total (Pockets 4-9); however, the amino acid residues of Pocket 4 at the positions β70, β71 and β74 (located in the *DQB* chain) are suggested to be of significant importance (Ou *et al.* 1998) as their topology is such that they protrude above the plane of the PBR and thus coming to direct contact to the T-cell receptor (TCR; Stern *et al.* 1994). These residues are among the most polymorphic residues of the PBR (Stern *et al.* 1994) and this is also observed in the cetacean alleles identified in this study (Figure 3.2.1). The amino acid side chain charges are key influence parameters for the interaction between TCR, antigen and MHC molecule (Ou *et al.* 1998).

Hill *et al.* (1991) showed that class I B*53 and class II *DRB1**1302 alleles are over-represented in individuals who are resistant to severe malaria and that also these alleles are significantly more frequent among the west African population than other racial groups. The west African population is exposed to malaria parasite (Hill *et al.* 1991). In addition, De Campos-Lima *et al.* (1993) have shown that a mutant strain of Epstein-Barr virus is associated to populations with high frequencies of class I *A*11 allele where it escapes presentation, while the wild type strain which is effectively presented by *A*11 allele, is found in populations where this allele is rare. Both of these studies suggested that pathogen driven selection shapes MHC polymorphism (Hill *et al.* 1991; De Campos-Lima *et al.* 1993; Meyer and Thomson 2001).

In the six focal species of cetaceans, the total charge of each of the allele was estimated and alleles were categorized into the four functional groups of positive (+), negative (-), di-charged (+/-) and neutral (n) charge accordingly (see Chapter 2: 2.6.1.4). The pattern of total charge distribution in the populations follows the population differentiation pattern according to MHC *Fst* analysis (see Figures 5.2.1-5.2.4, Tables 5.2.1-5.2.4(iii) and Figures 5.2.5-5.2.8). Support for the role of selection in *DQB1* genetic differentiation is invoked due to the fact that Fst *DQB1* analysis does not correlate to the one inferred by microsatellite DNA. In addition, the results from the nucleotide substitution analysis showed that *B. bonaerensis* population of A5 had significantly higher rates of non-synonymous substitutions than both populations of *B. acutorostrata*. Furthermore, the M10 allele, which is exclusive to the A5 population, is the only 'positively' charged allele identified in this study in minke whale populations. Differential selection across habitats may result in habitat specific allelic composition at the *DQB1* locus in cetaceans (Cohen 2002).

5.3.2 Conclusions

The population genetic analysis of the MHC *DQB1* locus provides further support to the hypothesis that MHC *DQB1* variation in cetacean species is maintained by selection and this locus is involved in cetacean immune response. This results from the deviations from Hardy-Weinberg equilibrium and heterozygote excess (high negative values of *Fis*) at population level, from *Fst* values, which are significant different to the values obtained by neutral markers (such as nuclear and mitochondrion DNA) and also from the structural analysis of the PBR region which follows the high

polymorphism at Pocket 4 amino acid residues pattern derived from human class II loci. In addition, this analysis may offer an advantage over the allelic frequency methods used in the population genetics analyses in its ability to suggest a functional relation between allele frequencies and environmental stressors. The study of MHC variation at a population level can be very informative for wildlife populations, not only for the purpose of how selection maintains variation in MHC but for the identification of population-specific environmental stressors which may have an implication in the conservation and management of wild populations.

CHAPTER 6:

GENERAL DISCUSSION

6.1 Evolution of the *DQB1* locus in cetaceans

The core hypothesis of the present study was that diversity in exon-2 MHC *DQB1* locus in cetaceans is generated and maintained by selection, through the marine pathogen environment. Previous studies in cetaceans suggested low levels of variation in *DQB1* locus (Trowsdale *et al.* 1989; Murray *et al.* 1995; 1998). However this study indicates that variation in this locus in cetaceans is comparable to terrestrial mammals. Combinations of phylogenetic and genetic analyses were used in order to assess the role of selection at the order, suborder, species and population level in an extensive sample of cetacean species. The hypothesis is supported at all different levels.

The polymorphism in the *DQB1* locus in cetaceans is consistent with established characteristics of MHC variation (*e.g.* Klein and Sato 1998; Meyers and Thompson 2000; Garrigan and Hedrick 2003). Allelic sequences identified in this study proved to be highly divergent with most nucleotide substitutions being non-synonymous (Chapter 3). Analyses in the order (Chapter 3), suborder (Chapter 4), species (Chapter 4) and population (Chapter 5) level showed that variable amino acids were predominantly at or near positions known to be of a functional importance in the *DQB1* (PBR sites). The phylogeny of the cetacean alleles at the order and suborder level attests to the trans-species evolution pattern of MHC polymorphism. Population significant departures from the expectations of the Hardy-Weinberg equilibrium and population differentiation (*Fst* analysis) according to *DQB1* locus support selection at the population level. There are three forms of balancing selection which have been invoked from previous studies on MHC polymorphism in terrestrial mammals: overdominant, frequency dependent and fluctuating selection. I suggest that these selection forms can also be invoked from the results of the present study for the following reasons.

The genetic structure of pathogen populations is shaped by recombination, mutation, competition from other pathogens and the immune system of the host (Hedrick *et al.* 1991). This determines the antigenic diversity. The same applies to

MHC genes as the immune response of the host has to evolve in order to respond to pathogen evolution. Thus MHC variation is maintained and determined through pathogen interaction.

Several studies have shown that there is an association between certain MHC alleles and specific pathogens, which supports the notion of direct selection on MHC alleles by pathogens (Hill *et al.* 1995; Thurz *et al.* 1995; Paterson *et al.* 1998; Godot *et al.* 2000; Langefors *et al.* 2001). The population diversity of MHC molecules is extremely large and certain HLA loci can have hundreds of alleles (Meyer and Thompson 2001). Due to the high degree of polymorphism, pathogens which escape presentation by one host may not in another. Thus, a heterozygous host will have an advantage over a homozygous, as it will be able to present a greater array of peptides (Doherty and Zinkernagel 1975; Klein 1986; Hughes and Nei 1988, 1989).

In Chapter 3, Chapter 4 and Chapter 5, it was shown that the ratio of non-synonymous to synonymous substitutions (*dn/ds*) in PBR is significantly higher than one and significantly higher than the non-PBR sites, supporting the expectations of balancing selection (Hughes and Nei 1988, 1989; Takahata and Nei 1990). Furthermore, the proportion of trans-species lineages observed in the phylogeny of *DQB1* alleles conforms with the expectations of balancing selection (Takahata and Nei 1990; Takahata *et al.* 1992). In Chapter 4, it was observed that there is a positive correlation between the number of median vectors (which denote extant unsampled alleles or extinct alleles; see Chapter 2: 2.6.2.5) and the *dn/ds* ratio value in the eight focal species (see Table 4.2 for statistic values). This observation provides further support for balancing selection in cetacean *DQB1*, as Takahata *et al.* (1992) demonstrated that there is a correlation between the allele number and the substitution differences in PBR.

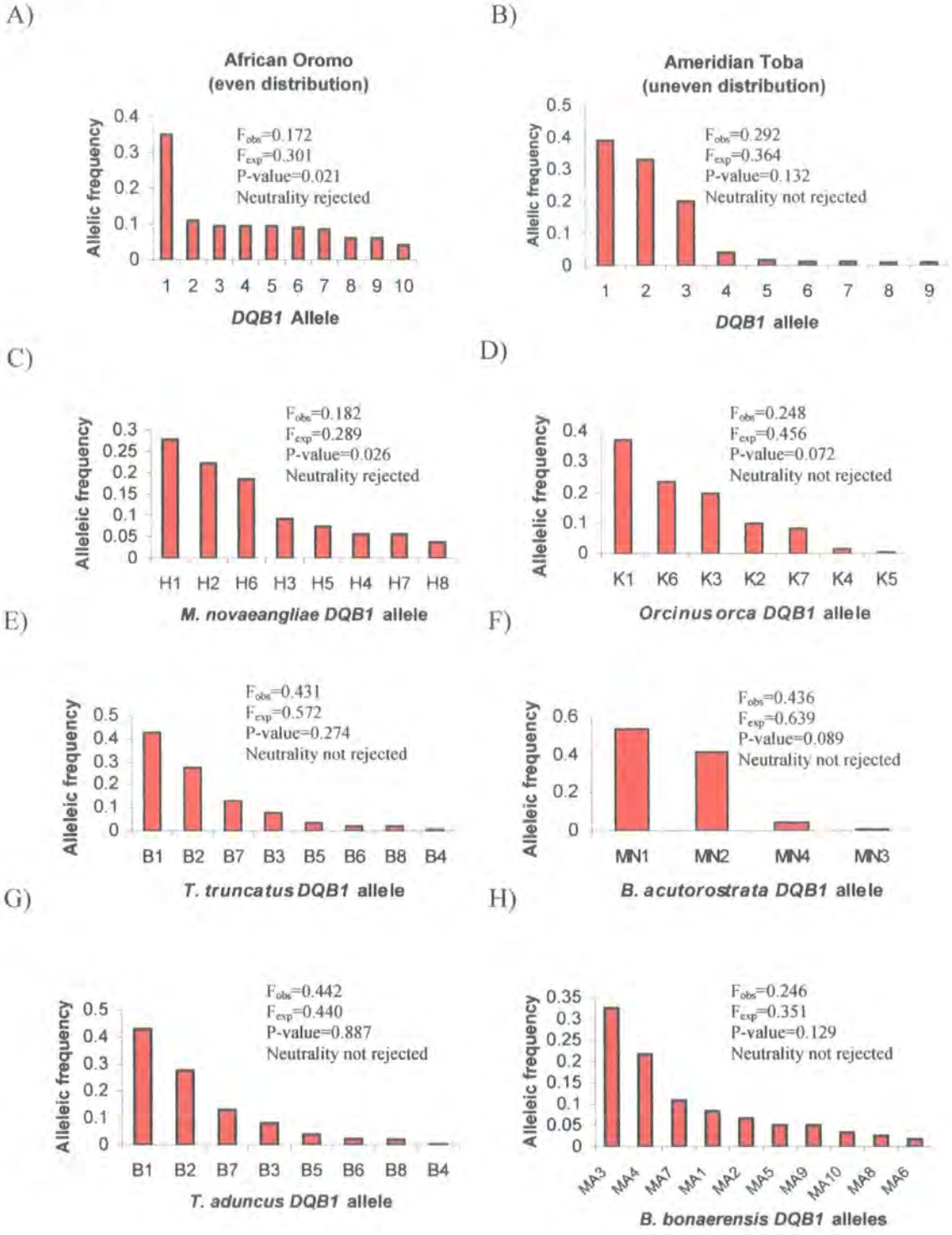
Heterozygote advantage (overdominant selection) is one of the main selection mechanisms proposed to explain the high observed level polymorphism in the MHC. The simplest form of overdominant selection is referred to as symmetric overdominance which assumes that all heterozygotes have higher and equal fitness while homozygotes have lower and equal fitness values (Lewontin *et al.* 1978). Kimura and Crow (1964) and Maruyama and Nei (1981) showed that overdominant selection increases the proportion of heterozygotes, heterozygosity and mean number of alleles at a given locus an observation which is consistent with the observed high levels of polymorphism in the MHC.

In several human populations, such as Saharan Africans, South American Amerindians and North American Amerindians, the frequency of homozygotes is significantly lower than that according to Hardy-Weinberg expectations, an observation attributed to overdominant selection (Degos *et al.* 1974; Black and Salzano 1981; Markow *et al.* 1993). This is also observed in Chapter 5, where 78% of the cetacean populations studied, deviated significantly from Hardy-Weinberg expectations and exhibit a homozygote deficiency, according to expectations of overdominant selection. Furthermore, Takahata *et al.* (1992) demonstrated that polymorphism generated by overdominant selection requires that new alleles are introduced into the population over time, allowing new fitness arrays to be created. This evolution pattern was observed in the network phylogenies of *Odontoceti* and *Mysticeti* species, as the torso of the network phylogeny comprises a high number of median vectors, which hypothetically supports the expectation of overdominant selection (Takahata *et al.* 1992).

However, the allelic frequencies observed in the population analyses do not exhibit an even distribution according to overdominant selection expectations (Figures 5.2.1–5.2.4; Meyer and Thompson 2001). Figure 6.1.1 A and B show the allelic frequencies distribution of *DQB1* locus in two human populations, the Brazilian Toba (Cerna *et al.* 1993) and the African Oromo (Fort *et al.* 1998). In the Oromo population, allelic frequencies significantly depart form neutrality expectations exhibiting an even distribution (consistent with balancing selection expectations). In the Toba population the allelic frequencies do not refute neutrality and therefore do not conform to the even distribution expectations of balancing selection (Cerna *et al.* 1993; Meyer and Thomson 2001). The cetacean populations also do not refute neutrality and allelic frequencies are not distributed evenly (Figure 6.1.1 C-H).

Although the significance of overdominant selection mechanism is undisputed, studies on MHC polymorphism have argued that this selection mechanism solely cannot justify the observed polymorphism exhibited by MHC (Lewontin *et al.* 1978; Aoki *et al.* 1980; Parham *et al.* 1989; Lawlor *et al.* 1990; Wills 1991). Frequency dependent selection (host-pathogen co-evolution) has also been invoked in order to explain MHC polymorphism.

Figure 6.1.1 (A-H): Shows the allelic frequency distribution in the African Oromo (Fort *et al.* 1997) and the Ameridian Toba (Cerna *et al.* 1993) and in the six focal cetacean species included in the present study.



Under frequency dependent selection, fitness values are not fixed but change in proportion to allelic frequencies (Meyer and Thomson 2001). For example, pathogen mutations, which escape presentation by frequent alleles, will be promoted by selection and the fitness of these alleles will decrease. However, the frequency of rare or novel alleles in the population will increase, since few pathogens have been exposed and adapted to. Furthermore, once the pathogens adapt to the rare or novel alleles, the fitness will decline, as pathogens adapting to these will be favored. This will lead into allele frequencies fluctuating over time as pathogens adapt to them. Trematode parasite in wild populations of snails has been shown to be under frequency-dependent selection, as it was demonstrated that the parasite adapted to be most virulent in the dominant genotype of the host (Dybdahl and Lively 1998; Lively and Dybdahl 2000). In addition, HIV-1 virus in humans reaches a higher viral load in individuals with common MHC molecules (Trachtenberg *et al.* 2003).

De Boer *et al.* (2004) argued that in order for overdominant selection to account for the observed high levels of polymorphism in MHC, different alleles have to confer very similar fitness, an expectation which is not supported by experimental data (Carrington *et al.* 1999, Jeffery *et al.* 2000; Penn *et al.* 2002; Acevedo-Whitehouse *et al.* 2003). Borghans *et al.* (2004) developed a computer simulation of coevolving hosts and pathogens and demonstrated that overdominant selection *per se* is not sufficient to explain the high level of polymorphism in MHC whilst frequency dependent selection, through host-pathogen coevolution, supports realistic MHC polymorphisms.

In the present study, the distribution of allelic frequencies observed in the population analysis may indicate frequency-dependent selection, since their distributions differ within and among populations (Chapter 5). According to the expectations of this type of selection, the allelic frequency fluctuates over time as pathogens adapt to the most frequent alleles (Takahata and Nei 1990; Meyer and Thomson 2001). In addition, a study on great reed warblers (*Acrocephalus arundinaceus*) by Westerdahl *et al.* (2004) suggested that MHC allele frequencies might be explained by fluctuating selection.

In fluctuating selection, the fitness of the population changes according to the frequency of the pathogens whilst in frequency-dependent selection the fitness of the population changes according to the frequency of the alleles (Hedrick et al. 1987; Meyer and Thomson 2001). Fluctuating selection has been suggested as an explanation of different allele frequencies among populations of the same species and has been shown

to create more variation in MHC alleles than in neutral alleles (Chapter 5 Discussion; Westerdahl *et al.* 2004).

MHC heterozygosity has been shown to correlate negatively with disease in HIV-infected patients (Carrington *et al.* 1999), with HTLV-1 infection in humans (Jeffery *et al.* 2000) and with infectious diseases in Californian sea lions (*Zalophus californianus*; Acevedo-Whitehouse *et al.* 2003). It has been shown that in eight natural populations of three-spined sticklebacks (*Gasterosteus aculeatus* L.), individuals which present a high number of MHC alleles were less susceptible to induced pathogens (Kurtz *et al.* 2004). McClelland *et al.* 2002 were the first to experimentally show that mice that were heterozygous at MHC loci were superior to either homozytes upon infection with multiple pathogens. However, Borghans *et al.* (2004) suggested that MHC polymorphism is the result of the dynamic equilibrium between overdominant selection and frequency dependent selection through host-pathogen co-evolution. The observed polymorphism of the *DQB1* locus in cetaceans may also be interpreted according to the expectations of these selective forces.

The role of selection in generating *DQB1* diversity in cetaceans was also assessed through the environmental context. The results of the present study also provide support for the hypotheses that differences in environmental parameters, such as social behavior, social structure and habitat differentiation, may affect pathogen selection pressures. In Chapter 4, it was shown that the rates of non-synonymous substitutions among different cetacean species were significantly different and in Chapter 5 it was shown that population differentiation according to *DQB1* was significantly different than that of microsatellite DNA. Furthermore, populations from different habitats and geographical regions exhibited differential distribution of alleles and allelic frequencies (see Chapter 5). Previous studies have suggested that differences in selection pressure result in population-specific MHC variation through balancing selection (see Chapter 5; Hill 1998; Landry and Bernatchez 2001; Hedrick 2002; Cohen 2003).

6.2 Limitations of the present study and recommendations for future research

There are three main limitations which are found in my study: i) MHC polymorphism evaluation was based on the analysis of the Peptide Binding Region of

class II exon-2 *DQB1* locus, ii) the sample size of *Mysticeti* species was relatively smaller than *Odontoceti* species (Chapter 4) and iii) 44% of the cetacean populations studied were comprised of 10-15 individuals (Chapter 5).

For the purpose of this study, genetic analyses were focused on exon-2 of *DQB1* locus, which codes for the PBR of the MHC molecule. The reasons for choosing *DQB1* locus are mentioned in Chapter 1, however, future analyses of further class II and class I genes would likely provide useful corroborating data.

The general aim of my study was to provide a further insight into the evolution and function of an MHC gene in cetaceans, in order to evaluate whether the marine environment has affected the immune response through MHC polymorphism. An extension of the present study could perhaps include a higher number of class I and class II MHC loci, including an assessment of nucleotide diversity in exons and introns.

Previous studies in humans have shown that there is a wide range of correlation between different loci and different racial groups and between different allelic frequencies and different populations, a fact that has been suggested to be the result of different selection pressures (Wills 1991). In addition, Sanchez-Mazas *et al.* (2000) have shown that all loci within the class II region (with the exception of DPB1) in a sample of 100 French individuals are in linkage disequilibrium with each other, an observation which has been attributed to selection (Sanchez-Mazas et al. 2000).

Class I loci have been identified in pinnipeds (Bowen *et al.* 2003) and cetaceans (Rodrigez *et al.* 1999) suggesting that class I molecules are also involved in marine mammal immune response (Rodrigez *et al.* 1999; Romano *et al.* 2001). An assessment of MHC polymorphism in class I molecules in cetaceans could provide further information on MHC evolution and antigen presentation in cetaceans.

The sample size of *Mysticeti* species included in this study was relatively small compared to *Odontoceti* species, which resulted in a higher number of alleles present in the *Odontoceti* sample. This may have an affect on the comparison of trans-species lineages and nucleotide substitution patterns between *Odontoceti* and *Mysticeti.*

The present study suggested that different populations exhibit differences in allele frequencies and substitution patterns (Chapter 5). However, a relatively small sample size may affect allelic distribution and their frequencies (Sokal and Rohlf 1995). Future work should include a more extensive sample size of different populations and also populations of different species of cetaceans which share the same habitat in order to evaluate possible differences in selection pressure.

Studies on mice and humans have shown that mate preference may also affect MHC polymorphism. Studies in mice have suggested that non-random mating in mice is correlated to MHC incompatibility (Egid and Brown 1989; Potts *et al.* 1991). In addition, Yamazaki *et al.* (1983) demonstrated that mice can distinguish between con-specifics that differ solely at MHC genes. Furthermore, Ober *et al.* (1997) showed that Hutterite (a Caucasian north American reproductive isolate of European ancestry) mate choice is influenced by HLA haplotypes, avoiding spouses with the same haplotype as their own. MHC recognition in mice (urine odors) and humans (urine-odors and sweat) is suggested to be determined through olfactory senses (Wedekind *et al.* 1995; Wedekind and Furi 1997). Certain populations of killer whale (British Columbia and Washington residents; Ford 2002) and bottlenose dolphin (west coast of Florida; Connor 1999) have been the subject of long-term studies, including extensive photo-identification of individuals and family trees. These populations offer the opportunity to assess mate preference in cetaceans and MHC polymorphism.

REFERENCES

Acevedo-Whitehouse, K., Gulland, F., Greig, D. and Amos, W. (2003) Inbreeding: disease susceptibility in California sea lions. *Nature*, **422**, 35-37.

Aguilar, A., Hammond, P.S., Pastor, X. and Aguilar, R. (1994) Distribution and numbers of striped dolphins in the western Mediterranean sea after the 1990 epizootic. *Marine Mammal Science*, **10**, 137-150.

Aguilar, A., Roemer, G., Debenham, S., Binns, M., Garcelon, D. and Wayne, R.K. (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences*, **101**, 3490-3494.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (2002) *Molecular Biology of the Cell*. Garland, London.

Anderson, E. (1995) Morbillivirus infections in wildlife (in relation to their population biology and disease control in domestic animals). *Veterinary Microbiology*, **44**, 319-332.

Anderson, M.J., Hessel, J.K. and Dixson, A.F. (2004) Primate mating systems and the evolution of immune response. *Journal of Reproductive Immunology*, **61**, 31-38.

Archer, F.I. and Perrin, W.F. (1999) *Stenella coeruleoalba. Mammalian Species*, **603**, 1-9.

Avise, J.C. (1994) *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Bacon, P.J. (ed.). (1985) *Population dynamics of rabies in the wildlife*. Academic Press Ltd, London.

Baird, R.W. (2001) Status of killer whales in Canada. *Canadian Field-Naturalist*, **28**, 45-59.

Baird, R.W. (2002) Risso's dolphin. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Bandelt, H.-J., Forster, P. and Rohl, A. (1999) Median -joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37-48.

Bannister, J.L. (2002) Baleen Whales. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Barrett, T., Visser, I.K., Mamaev, L. and Goatley, L. (1993) Dolphin and porpoise morbilliviruses are genetically distinct from phocine distemper virus. *Veterinary Medicine*, **193**, 1010-1012.

Barrett, T., Blixenkrone-Moller, M., Di Guardo, G., Domingo, M., Duignan, P., Hall, A., Mamaev, L. and Osterhaus, A.D.M.E. (1995) Morbilliviruses in aquatic mammals: report on round table discussion. *Veterinary Microbiology*, **44**, 261-265.

Barrett, T. (1999) Morbillivirus infections, with special emphasis on morbilliviruses of carnivores. *Veterinary Microbiology*, **69**, 3-13.

Barrett-Lennard, L.G. (2000) Population structure and mating patterns of killer whales (*Orcinus orca*) as revealed by DNA analysis. University of British Columbia, Vancouver, BC.

Beardmore, I. and White, K.A.J. (2001) Spreading disease through social groupings in competition. *Journal of Theoretical Biology*, **212**, 253-269.

Becker, P.R., Mackey, E.A., Demiralp, R., Schantz, M.M., Koster, B.J. and Wise, S.A. (1997) Concentrations of chlorinated hydrocarbons and trace elements in marine mammal tissues archived in the U.S. national biomonitoring specimen bank. *Chemosphere*, **34**, 2067-2098.

Becker, P.R. (2000) Concentration of chlorinated hydrocarbons and heavy metals in Alaska arctic marine mammals. *Marine Pollution Bulletin*, **40**, 819-829.

Begon, M., Bennet, M., Bowers, R.G., French, N.P., Hazel, S.M. and Turner, J. (2002) A clarification of transmission terms in host-microparasite models; numbers, densitites and areas. *Epidemiology and Infection*, **129**, 147-153.

Berta, A. and Sumich, J.L. (1999) *Marine Mammals Evolutionary Biology*. Academic Press, USA, San Diego.

Bigg, M.A., Olesiuk, P.F., Ellis, G.M. and Ford, J.B.K. (1990) Social organization and genealogy of resident killer whale (*Orcinus orca*) in the coastal waters of British Columbia and Washington state. *Reprint International Whaling Commission Special Issue*, **12**, 383-405.

Binh Minh, T., Watanabe, M., Nakata, H., Tanabe, S. and Jefferson, T.A. (1999) Contamination by persistent organochlorines in small cetaceans from Hong Kong coastal waters. *Marine Pollution Bulletin*, **39**, 383-392.

Birkun, A., Kuiken, T., Krivokhizhin, S., Haines, D.M., Osterhaus, A.D.M.E., van de Bilt, M.W.G., Joiris, C.R. and Siebert, U. (1999) Epizootic of morbilliviral disease in common dolphins (*Delphinus delphis ponticus*) from the Black sea. *Veterinary Record*, **144**, 85-92.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction proceedure for screening

recombinant plasmid DNA. *Nucleic Acids Research*, **24**, 1513-1523.

Black, F.L. and Salzano, F.M. (1981) Evidence for heterosis in the HLA system. *American Journal of Human Genetics*, **33**, 894-899.

Bodmer, W.F. (1972) Evolutionary significance of the HLA system. *Nature*, **237**, 139-145.

Bolt, G. and Blixenkrone-Moller, M. (1994) Nucleic acid hybridization analyses confirm the presence of a hitherto unknown morbillivirus in Mediterranean dolphins. *Veterinary Microbiology*, **41**, 363-372.

Borghans, J.A.M., Beltman, J.B. and deBoer, R.J. (2004) MHC polymorphism under host-pathogen coevolution. *Immunogenetics*, **55**, 732-739.

Bowers, R.G. and Turner, J. (1997) Community structures and the interplay between interspecific infection and competition. *Journal of Theoretical Biology*, **187**, 95-107.

Boyce, W.M., Hedrick, P.W., Muggli-Cockett, N.E., Kalinowski, S., Penedo, M.C.T. and Beyll, R.R. (1997) Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in Bighorn sheep. *Genetics*, **145**, 421-433.

Bronwell, R.L., Curry, B., Van Bonn, W. and Ridgay, S.H. (2000) Conservation Conundrum. *Science*, **288**, 2319-2320.

Brown, J.H.T., Jardetzky, M.A., Saper, B., Samraoui, P., Bjorkman, J. and Wiley, D.C. (1988) A hypothetical model of the foreign antigen binding site of Class II histocompatibility molecules. *Nature*, **332**, 845-850.

Brownell, R.L., Perrin, W.F., Pastene, L.A., Palsboll, P.J., Mead, J.G., Zerbini, A.N., Kasuya, T. and Tormosov, D.D. (2000) Worldwide taxonomic status and geographic distribution of minke whales (*Balaenoptera acutorostrata* and *B. bonaerensis*). *International Whaling Commission meet. doc. SC/52/OS3/O27, available from IWC, Cambridge, UK*, 1-13.

Carrington, M., Nelson, G.W., Martin, M.P., Kissner, T., Vlahov, D., Goedert, J.L., Kaslow, R., Buchbinder, S., Hoots, K. and O'Brien, S.J. (1999) HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science*, **283**, 1748-1752.

Carwardine, M. (1995) *Whales, dolphins and porpoises*. Dorling Kindersley, London.

Cerna, M., Falco, M., Friedman, H., Raimondi, E., Maccagno, A., Fernandez-Vina, M. and Stastny, P. (1993) Differences in HLA Class II alleles of isolated South American Indian populations form Brazil and Argentina. *Human Immunology*, **37**, 213-220.

Chardon, P., Renard, C. and Vaiman, M. (1999) The major histocompatibility complex in swine. *Immunological Reviews*, **167**, 179-192.

Charles, L.N., Gittleman, J.L. and Janis, A. (2000) Promiscuity and the Primate Immune System. *Science*, **290**, 1168-1170.

Cheesman, C.L. and Mallison, P.J. (1982) Behavior of badgers (*Meles meles*) infected with bovine tuberculosis. *Journal of Zoology*, **194**, 284-289.

Chivers, S.J. (2002) Cetacean life history. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Chung, C.T. and Miller, R.H. (1988) A rapid and convenient method for preparation and storage of competent bacterial cells. *Nucleic Acids Research*, **25**, 3580-3581.

Clapman, P.J. (1996) The social and reproductive biology of humpback whales; an ecological perspective. *Mammal Reviews*, **26**, 27-49.

Clapman, P.J. (2002) Humpback whale. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Cliff, A.D. (1995) Incorporating spatial components into models of epidemic models. In Mollison, D. (ed.), *Epidemic models-their structure and relation to data*. Cambridge University Press, Cambridge.

Cohen, S. (2002) Strong positive selection and habitat-specific amino acid substitution patterns in *Mhc* from an estuarine fish under intense pollution stress. *Molecular Biology and Evolution*, **19**, 1870-1880.

Connor, R.C., Mann, J., Tyack, P.L. and Whitehead, H. (1998) Social evolution in toothed whales. *Trends in Ecology and Evolution*, **13**, 228-232.

Connor, R.C., Wells, R.S., Mann, J. and read, A.J. (1999) The bottlenose dolphin, *Tursiops* spp: Social relationships in a fission-fusion society. In Mann, J., Connor, R.C., Tyack, P.L. and Whitehead, H. (eds.), *Cetacean societies: field studies of dolphins and whales*. The University of Chicago Press, Chicago.

Connor, R.C. (2002) Ecology of group living and social behaviour. In Hoelzel, A.R. (ed.), *Marine mammal biology; an evolutionary approach*. Blackwell Science Ltd, Oxford.

Dahleim, M.E. and Heyning, J.E. (1999) Killer Whale *Orcinus orca* (Linneus 1758). In Ridgway, S.H. and Harrison, R. (eds.), *Handbook of Marine Mammals*. Academic Press Ltd, San Diego.

Day, T. and Burns, J.G. (2003) A consideration of patterns of virulence arising from host-parasite coevolution. *Evolution*, **57**, 671-676.

De Boer, R.J., Borghans, J.A.M., Van Bowen, M., Kesmir, C. and Weissing, F.J. (2004) Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics*, **55**, 725-731.

De Campos-Lima, P.O., Gavioli, R., Zhang, Q.J., Wallace, L.E., Dolcetti, R., Rowe, M.,

Rickinson, A.B. and Masucci, M.G. (1993) HLA-A11 epitope loss isolates of Epstein-Barr virus from a highly A11+ population. *Science*, **260**, 98-100.

De Guise, S., Bernier, J., Dufresne, M.M., Martineau, D., Beland, P. and Fournier, M. (1996) Immune functions in beluga whales (*Delphinapterus leucas*): evaluation of mitogen-induced blastic transformation of lymphocytes from peripheral blood, spleen and thymus. *Veterinary Immunology and Immunopathology*, **50**, 117-126.

De Guise, S., Ross, P.S., Osterhaus, A.D.M.E., Martineau, D., Beland, P., Fournier, M. (1997) Immune functions in beluga whales (*Delphinapterus leucas*): evaluation of natural killer cell activity. *Veterinary Immunology and Immunopathology*, **58**, 345-354.

De Muisson, C. (2001) Walking with whales. *Nature*, **413**, 259-260.

Degos, L., Colombani, J., Chaventre, A., Bengston, B. and Jacquard, A. (1974) Selective pressures on HLA polymorphism. *Nature*, **249**, 62-63.

Di Guardo, G., Agrimi, U., Amaddeo, D., McAliskey, M. and Kennedy, S. (1992) Morbillivirus infection in a striped dolphin (*Stenella coeruleoalba*) from the coasts of Italy. *Veterinary Record*, **130**, 579-580.

Doherty, P.C. and Zinkernagel, R.M. (1975) Enhanced immunological surveillance in mice heterozygous at the *H-2* gene complex. *Nature*, **256**, 50-52.

Domingo, M., Vilafranca, M., Visa, J., Prats, N., Trudgett, A. and Visser, I. (1995) Evidence for chronic morbillivirus infection in the Mediterranean striped dolphin (*Stenella coeruLeoalba*). *Veterinary Microbiology*, **44**, 229-239.

Du Pasquier, L. (1992) Origin and evolution of the vertebrate immune response. *APMIS*, **100**, 383-392.

Duignan, P.J., House, C., Geraci, J.R., Early, G., Copland, H.G., Walsh, M.T., Bossart, G.D., Cray, C., Sadove, S., Aubin, D.J.S. and Moore, M. (1995) Morbillivirus infection in two species of pilot whales (*Globicephala* Sp) from the Western Atlantic. *Marine Mammal Science*, **11**, 150-162.

Duignan, P.J., House, C., Geraci, J.R., Duffy, N., Rima, B.K., Walsh, M.T., Early, G., Aubin, D.J.St., Sadove, S., Koopman, H., Rhinehart, H. (1995) Morbillivirus infection in cetaceans of the Western Atlantic. *Veterinary Microbiology*, **44**, 241-249.

Duignan, P.J., Nielsen, O., House, C., Kovacs. K,M., Duffy, N., Early, G., Sadove, S., Aubin, D.J.St., Rima, B.K., Geraci, J.R. (1997) Epizootiology of morbillivirus infection in Harp, Hooded and ringed seals from the Canadian Arctic and Western Atlantic. *Journal of Wildlife Diseases*, **33**, 7-19.

Dybdahl, M.F. and Lively, C.M. (1998) Host-parasite coevolution: evidence for rare advantage and time-lagged selection in a natural population. *Evolution*, **52**,

1057-1066.

Edwards, S.V. and Hedrick, P.W. (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology and Evolution*, **13**, 305-311.

Egit, K. and Brown, J.L. (1989) The major histocompatibility complex and female mating preferences in mice. *Nature*, **351**, 548-550.

Ellegren, H., Hartman, G., Johansson, M. and Andersson, L. (1993) Major histocompatibility complex monomorphism and low levels of DNA fingerprinting variability in a reintroduced and rapidly expanding population of beavers. *Proceedings of the National Academy of Sciences*, **90**, 8150-8153.

Ellegren, H., Mikko, S., Wallin, K. and Andersson, L. (1996) Limited polymorphism at major histocompatibility complex (MHC) loci in the Swedish moose *A. alces*. *Molecular Ecology*, **5**, 3-9.

Ellis, S.A., Holmes, E.C., Staines, K.A., Smith, K.B., Stear, M.J., McKeever, D.J., MacHugh, N.D. and Morrison, W.I. (1999) Variation in the number of expressed MHC genes in different cattle class I haplotypes. *Immunogenetics*, **50**, 319-328.

Erlich, H.A., Gelfand, D. and Sninsky, J.J. (1991) Recent Advances in the Polymerase Chain Reaction. *Science*, **252**, 1643-1651.

Evans, P.G. (1987) *The Natural History of Whales and Dolphins*. Christopher Helm Ltd., London.

Evans, W.E. (1994) Common dolphin white-bellied porpoise *Delphinus delphis* Linneus, 1758. In Ridgway, S.H. and Harrison, R. (eds.), *Handbook of Marine Mammals*, Vol. 5.

Evans, P.G.H. and Raga, J.A. (eds.). (2001) *Marine Mammals; biology and conservation*. Kluwer Academic/ Plenum Publishers, New York.

Felsenstein, J. (1981) Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution*, **35**, 1229-1242.

Felsenstein, J. (1988) Phylogenis from molecular sequences: inference and reliability. *Annual Review in Genetics*, **22**, 521-565.

Fernandez-Vina, M., Lazaro, A.M., Marcos, C.Y., Nulf, C., Raimondi, E., Hass, E.J. and Stastny, P. (1997) Dissimilar evolution of B-locus versus A-locus and class II loci of the HLA region in South American Indian tribes. *Tissue Antigens*, **50**, 233-250.

Fertl, D. (2002) Fisheries, interference with. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Fitch, W.M. (1997) Networks and viral evolution. *Molecular Biology and Evolution*, **44**,

S65-S75.

Flores-Ramirez, S., Urban-Ramirez, J. and Miller, R.D. (2000) Major Histocompatibilty Complex Class I Loci from the Gray Whale (*Eschrichtius robustus*). *The Journal of Heredity*, **91**, 279-282.

Ford, J.B.K., Ellis, G.M., Barrett-Lennard, L.G., Morton, A.B., Palm, R.S. and Balcomb III, K.C. (1998) Dietary specialization in two sympatric populations of killer whales (*O. orca*) in coastal British Columbia and adjacent waters. *Canadian Journal of Zoology*, **76**, 1456-1471.

Ford, J.B.K. and Ellis, G.M. (1999) *'Transients: mammal hunting killer whales of British Columbia, Washington and Southeastern Alaska.* UBC Press and University of Washington Press, Vancouver, BC and Seattle, WA.

Ford, B.K. (2002) Killer whale. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals.* Academic Press, San Diego.

Forsyth, M.A., Kennedy, S., Wilson, S., Eybatov, T. and Barrett, T. (1998) Canine distemper virus in a Caspian seal. *Veterinary Record*, **143**, 662-664.

Fort, M., deStefano, G.-F., Giraldo-Alvarez, P., Dugoujon, J.-M., Ohayon, E., Scano, G. and Abbal, M. (1997) HLA class II allele and haplotype frequencies in Ethiopian Amhara and Oromo populations. *Tissue Antigens*, **51**, 327-336.

Fu, X.-T., Bono, C.P., Woulfe, S.L., Swearingen, C., Summers, N.L., Sinigaglia, F., Sette, A., Schwartz, B.D. and Karr, R.W. (1995) Pocket 4 of the HLA-DR (a,b*0401) molecule is a major determinant of T cell recognition of peptide. *Journal of Experimental Medicine*, **181**, 915-926.

Garber, T.L., Hughes, A.L., Letvin, N.L., Templeton, J.W. and Watkins, D.I. (1993) Sequence and evolution of cattle MHC class I cDNAs: concerted evolution has not taken place in cattle. *Immunogenetics*, **38**, 11-20.

Garrigan, D. and Hedrick, P.W. (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *The Society for the Study of Evolution*, **57**, 1707-1722.

Gelberg, H.B., Dieterich, A. and Lewis, R.M. (1982) Vescicular Exanthema of Swine and San Miguel Sea Lion Virus: Experimental and Field Studies in Otarid Seals, Feeding Trials in Swine. *Veterinary Pathology*, **19**, 413-423.

Gillespie, J.H. (1998) *Population Genetics; a concise guide.* The Johns Hopkins University Press, Baltimore and London.

Gingerich, P.D., Haq, M.-u., Zalmout, I.S., Khan, I.H. and Malkani, M.S. (2001) Origin of whales form early artiodactyls: hands and feet of Eocene Protocetidae from Pakistan. *Science*, **293**, 2239-2242.

Gingerich, P.D. (2003) Land-to-sea transition in early whales: evolution of Eocene

Archaeoceti (*Cetacea*) in relation to skeletal proportions and locomotion of living semiaquatic mammals. *Paleobiology*, **29**, 429-454.

Goldman, N. and Yang, Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, **11**, 725-736.

Goudet, J. (2001) FSTAT: a program to estimate and test gene diversities and fixation indices. Available from: http:/ www. unil.ch/izea/softwares/fstat.html.

Goyal, S.M., Adams, W.N., O'Malley, M.L. and Lear, D.W. (1983) Human pathogenic viruses at sewage sludge disposal sites in the Middle Atlantic region. *Applied and Environmental Microbiology*, **48**, 758-763.

Gregorious, H.R. (1987) The relationship between the concepts of genetic diversity and differentiation. *Theoretical Applied Genetics*, **74**, 397-401.

Griffin, D.W., Donaldson, K.A., Paul, J.H. and Rose, J.B. (2003) Pathogenic human viruses in coastal waters. *Clinical Microbiology Reviews*, **16**, 129-143.

Gu, X. and Zhang, J. (1997) A simple method for estimating the parameter of substitution rate variation among sites. *Molecular Biology and Evolution*, **14**, 1106-1113.

Guardo, G.D., Agrimi, U., Morelli, L., Cardeti, G., Terracciano, G. and Kennedy, S. (1995) Post mortem investigations on cetaceans found stranded on the coasts of Italy between 1990 and 1993. *Veterinary Record*, **136**, 439-442.

Gudelj, I. (2004) The Effects of Spatial Movement and Group Interactions on Disease Dynamics of Social Animals. *Bulletin of Mathematical Biology*, **66**, 91-108.

Gudelj, I. and White, K.A.J. (2004) Spatial heterogeneity, social structure and disease dynamics of animal populations. *Theoretical Populations Biology*, **66**, 139-149.

Gyllensten, U.B. and Erlich, H.A. (1989) Ancient roots for polymorphism at the HLA-DQA locus in primates. *Proceedings of the National Academy of Sciences*, **86**, 9986-9990.

Gyllensten, U.B., Lashkari, D. and Erlich, H.A. (1990) Allelic diversification at the class II *DQB* locus of the mammalian major histocompatibility complex. *Proceedings of the National Academy of Sciences*, **87**, 1835-1839.

Haig, D. (1997) Maternal-Fetal interactions and MHC polymorphism. *Journal of Reproductive Immunology*, **35**, 101-109.

Hall, A.J., Law, R.J., Harwood, J., Ross, H.M., Kennedy, S., Allchin, C.R., Campbell, L.A. and Pomeroy, P.P. (1992) Organochlorine levels in common seal (Phoca vitulina) which were victims and survivors of the 1988 phocine distemper epizootic. *Science of the Total Environment*, **115**, 145-162.

Hamilton, W.D. (1971) Selection of selfish and altruistic behavior in some extreme

models. In Eisenberg and Dillon (eds.), *Man and Beast: comparative social behavior*. Smithsonian Institute Press, Washington, D. C.

Hammer, J., Gallazzi, F., Bono, E., Karr, R.W., Guenot, J., Valsasnini, P., Nagy, Z.A. and Sinigaglia, F. (1995) Peptide binding specificity of HLA-DR4 molecules: correlation with rheumatoid athritis association. *Journal of Experimental Medicine*, **181**, 1847-1855.

Harder, T.C., Plotz, J. and Liess, B. (1991) Antibodies against european phocine herpesvirus isolates detected in sera of Antarctic seals. *Polar Biology*, **11**, 509-512.

Hayashi, K., Nishida, S., Yosida, H., Goto, M., Pastene, L.A. and Koike, H. (2003) Sequence variation of the *DQB* allele in the cetacean MHC. *Mammal Study*, **28**, 89-96.

Hedrick, P.W., Klitz, W., Robinson, W.P., Kuhner, M.K. and Thomson, G. (1991) Evolutionary genetics of HLA. In Sclander, R., Clark, A.G. and Whittam, T. (eds.), *Evolution at the molecular level*. Sinnauer Association, Sunderland, MA.

Hedrick, P.W. (2002) Pathogen ressistance and genetic variation at MHC loci. *Evolution*, **56**, 1902-1908.

Heide-Jorgensen, M.P., Harkonen, T., Dietz, R. and Thompson, P.M. (1992) Retrospective of the 1988 European seal epizootic. *Diseases of Aquatic Organisms*, **13**, 37-62.

Heyning, J.E. and Lento, G.M. (2002) The evolution of marine mammals. In Hoelzel, A.R. (ed.), *Marine mammal biology; an evolutionary approach*. Blackwell Science Ltd, Oxford.

Higgins, R. (2000) Bacteria and fungi of marine mammals: A review. *Journal of Canadian Veterinary science*, **41**, 105-116.

Hill, A.V., Allsopp, C.E., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Bennett, P.M., Brewster, D., McMichael, A.J. and Greenwood, B.M. (1991) Common west African HLA antigens are associated with protection from severe malaria. *Nature*, **335**, 595-600.

Hill, A.V.S. (1998) The immunogenetics of human infectious diseases. *Annual Reviews in Immunology*, **16**, 593-617.

Hoelzel, A.R. and Dover, G.A. (1991) Genetic differentiation between sympatroc killer whale populations. *Heredity*, **66**, 191-196.

Hoelzel, A.R. and Dover, G.A. (1991) Mitochondrial D-loop DNA variation within and between populations of the minke whale (*Balaenoptera acutorostrata*). In Hoelzel, A.R. (ed.), *Genetic ecology of whales and dolphins*. International Whaling Commission, Cambridge.

Hoelzel, A.R., Potter, C.W. and Best, P.B. (1998) Genetic differentiation between parapatric 'nearshore' and 'offshore' populations of the bottlenose dolphin. *Proceedings of the National Academy of Sciences*, **265**, 1177-1183.

Hoelzel, A.R., Dahleim, M.E. and Stern, J.L. (1998) Low genetic variation among killer whales (*Orcinus orca*) in the eastern North Pacific, and genetic differentiation between foraging specialists. *Journal of Heredity*, **89**, 121-128.

Hoelzel, A.R. (ed.). (1998) *Molecular Genetic Analysis of Populations*. Oxford University Press.

Hoelzel, A. (1999) Impact of population bottlenecks on genetic variation and the importance of life-history; a case study of the northern elephant seal. *Biological Journal of the Linnean Society*, **68**, 23-29.

Hoelzel, A.R., Stephens, J.C. and O'Brien, S.J. (1999) Molecular Genetic Diversity and Evolution at the MHC DQB Locus in Four Species of Pinnipeds. *Molecular Biology and Evolution*, **16**, 611-618.

Hoelzel, A.R. (ed.). (2002) *Marine Mammal Biology; an evolutionary approach*. Blackwell Science Ltd.

Hoelzel, A.R., Natoli, A., Dahlheim, M.E., Olavarria, C., Baird, R.W. and Black, N.A. (2002) Low worldwide genetic diversity in the killer whale (*Orcinus orca*): implications for demographic history. *Proceedings of the Royal Society of London Biological Sciences*, **269**, 1467-1473.

Hoelzel, A.R., Goldsworthy, S.D. and Fleischer, R.C. (2002) Population genetic structure. In Hoelzel, A.R. (ed.), *Marine mammal biology; an evolutionary approach*. Blackwell Science Ltd, Oxford.

Holmes, E.C., Worebey, M. and Rambaut, A. (1999) Phylogenetic evidence for recombination in Dengue virus. *Molecular Biology and Evolution*, **16**, 405-409.

Holmes, E.C. and Ellis, S.A. (1999) Evolutionary history of MHC class I genes in the mammalian order *Perissodactyla*. *Journal of Molecular Evolution*, **49**, 316-324.

Holmes, E.C., Roberts, A.F.C., Staines, K.A. and Ellis, S.A. (2003) Evolution of major histocompatibility complex class I genes in Certartiodactyls. *Immunogenetics*, **55**, 193-202.

Horin, P. and Matiasovic, J. (2002) A second locus and new alleles in the major histocompatibility complex class I and class II (ELA) region in horses. *Animal Genetics*, **33**, 196-200.

Howard, E.B., Britt, J.O. and Matsumoto, G. (1983) Parasitic diseases. In Howard, E.B. (ed.), *Pathobiology of marine mammal diseases*. CRC Press, Florida, Vol. 1.

Hughes, A.L. and Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*,

335, 167-170.

Hughes, A.L. and Nei, M. (1989) Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proceedings of the National Academy of Sciences*, **86**, 958-962.

Hughes, A.L. and Nei, M. (1989) Evolution of the Major Histocompatibility Complex:independent origin of nonclassical class I genes in different groups of mammals. *Molecular Biology and Evolution*, **6**, 559-579.

Hughes, A.L. and Nei, M. (1990) Evolutionary relationships of class II Major-Histocompatibility-Complex genes in mammals. *Molecular Biology and Evolution*, **7**, 491-514.

Hughes, A.L., Ota, T. and Nei, M. (1990) Positive Darwinian selection promotes charge profile diversity in the antigen binding cleft of class I MHC molecules. *Molecular Biology and Evolution*, **7**, 515-524.

Hughes, A.L. (1991) MHC polymorphism and the design of captive breeding systems. *Conservation Biology*, **5**, 249-251.

Hughes, L.A. and Nei, M. (1992) Models of Host-Parasite Interaction and MHC polymorphism. *Genetics*, **132**, 863-864.

Hughes, A.L., Yeager, M. and Carrington, M. (1996) Peptide binding function and the paradox of HLA disease associations. *Immunology and Cell Biology*.

Hughes, A.L. and Yeager, M. (1997) Comparative evolutionary rates of introns and exons in murine rodents. *Journal of Molecular Evolution*, **45**, 125-130.

Hughes, A.L. and Yeager, M. (1998) Natural selection at Major Histocompatibility Complex loci of vertebrates. *Annual Review in Genetics*, **32**, 415-435.

Hughes, A.L. (1999) *Adaptive evolution of genes and genomes*. Oxford University Press, Oxford.

Huson, D.H. (1998) SplitsTree: analysing and visualizing evolutionary data. *Bioinformatics*, **14**, 68-73.

Huson, D.H. and Bryant, D. (2004) SplitsTree: a framework for building phylogenetic trees and networks.

Jaramillo, C., Montana, M.F., Castro, L.R., Vallejo, G.A. and Guhl, F. (2001) Differentiation and genetic analysis of *Rhodnius prolixus* and *Rhodnius colombiensis* by rDNA and RAPD amplification. *Mem Inst Oswaldo Cruz*, **96**, 1043-1048.

Jauniaux, T., Charlier, G., Desmecht, M. and Coignoul, F. (1998) Lesions of morbillivirus infection in a fin whale (*Balaenoptera physalus*) stranded along the Belgian coast. *Veterinary Record*, **143**, 423-424.

Jeffery, K.J., Siddiqui, A.A., Bunce, M., Lloyd, A.L., Vine, A.M., Witkover, A.D., Izumo, S., Usuku, K., Welsh, K.I., Osame, M. and R., B.C. (2000) the influence of HLA class I and heterozygosity on the outcome of human T cell lympotropic virus type I infection. *Journal of Immunology*, **165**, 7278-7284.

JIn, L. and Nei, M. (1990) Limitations of the evolutionary parsimony method of phylogenetic analysis. *Molecular Biology and Evolution*, **7**, 82-102.

Kaplan, C. (ed.). (1977) *Rabies: the facts*. Oxford University Press, Oxford.

Kennedy, S. (1990) A review of the 1988 European seal morbillivirus epizootic. *Veterinary Record*, **127**, 563-567.

Kennedy, S., Lindstedt, I.J., McAliskey, M.M., McConnel, S.A. and McCullough, S.J. (1992) Herpesviral encephalitis in a Harbor porpoise (*Phocoena phocoena*). *Journal of Zoo and Wildlife Medicine*, **23**, 374-379.

Kennedy, S., Kuiken, T., Ross, H.M., McAliskey, M., Moffett, D., McNiven, C.M., Carole, M. (1992) Morbillivirus infection in two common porpoises (*Phocoena phocoena*) from the coasts of England and Scotland. *Veterinary Record*, **131**, 286-290.

Kimura, M. and Crow, J.F. (1964) The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725-738.

Kimura, M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111-120.

Kimura, M. (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, Massachusetts.

Klein, J. (1978) H-2 mutations: their genetics and effect on immune function. *Journal of Advanced Immunology*, **26**, 55-146.

Klein, J. (1980) Generation of diversity to MHC loci: implications for T-cell receptor repertoires. In Fougereau, M. and Dausset, J. (eds.), *Immunology 80*. Academic Press Ltd, London.

Klein, J. (1986) *Natural History of the Major Histocompatibility Complex*. John Wiley and Sons, Inc.

Klein, J. and Figueroa, F. (1986) Evolution of the major histocompatibility complex. *Critical Reviews in Immunology*, **6**, 295-389.

Klein, J. (1987) Origin of Major Histocompatibility Complex: the trans-species hypothesis. *Human Immunology*, **19**, 155-162.

Klein, J. (1991) Of HLA, tryps and selection: an essay on coevolution of MHC and

parasites. *Human Immunology*, **30**, 247-258.

Klein, J. and O'hUigin, C. (1993) Composite origin of major histocompatibility complex genes. *Current Opinions in Genetic Development*, **3**, 923-930.

Klein, J. and O'hUigin, C. (1994) MHC polymorphism and parasites. *Proceedings of the Royal Society of London Biological Sciences*, **346**, 351-358.

Klein, J., Sato, A., Nagl, S. and O'hUigin, C. (1998) Molecular trans-species polymorphism. *Annual Review in Ecology and Systematics*, **29**, 1-21.

Klein, J. and Sato, A. (1998) Birth of the Major Histocompatibility Complex. *Scandinavian Journal of Immunology*, **47**, 199-209.

Kondo, R., Horai, S., Satta, Y. and Takahata, N. (1993) Evolution of hominoid mitochondrial DNA with special reference to the silent substitution rate over the genome. *Journal of Molecular Evolution*, **36**, 517-531.

Kosakovsky Pond, S.L. and Muse, S. (2004) Site-to-site variation of synonymous substitutions rates. *Genetics*, **22**, 34-45.

Kuby, J. (1994) *Immunology*. W. H. Freeman, New York.

Kumar, S., Tamura, K. and Nei, M. (1993) MEGA: molecular evolutionary genetic analysis. The Pennsylvania State University, Pennsylvania.

Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. (2001) MEGA: Molecular evolutionary genetics analysis. The Pennsylvania State University, Pennsylvania.

Kurtz, J., Kalbe, M., Aeschlimann, B.B., Haberli, M.A., Wegner, K.M., Reusch, T.B.H. and Milinski, M. (2004) Major histocompatibility complex influences parasite resistance and innate immunity in sticklebacks. *Proceedings of the Royal Society of London Biological Sciences*, **27**, 197-204.

Landry, C. and Bernatcez, L. (2001) Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*salmo salar*). *Molecular Ecology*, **10**, 2525-2539.

Lawlor, D.A., Ward, F.F., Ennis, P.D., Jackson, A.P. and Parham, P. (1988) HLA-A,-B polymorphisms predate the divergence of humans and chimpanzees. *Nature*, **335**, 268-271.

Lawlor, D.A., Zemmour, J., Ennis, P.D. and Prarham, P. (1990) Evolution of class-I MHC genes and proteins: from natural selection to thymic selection. *Annual Review of Immunology*, **8**, 23-37.

LeDuc, R. (2002) Delphinids, overview. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Levin, B.R., Lipsith, M. and Bonhoeffer, S. (1999) Population Biology, Evolution, and Infectious Disease: Convergence and Synthesis. *Science*, **283**, 806-809.

Lewontin, R., Ginzburg, L. and Tuljapurkar, S. (1978) Heterosis as an explanation for large amounts of genic polymorphism. *Genetics*, **88**, 149-170.

Li, W.-H., Wu, C.-I. and Luo, C.-C. (1985) A new method for estimating synonymous and non-synonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution*, **2**, 150-174.

Li, W.-H. (1993) Unbiased estimation of the rates of synonymous and non-synonymous substitution. *Journal of Molecular Evolution*, **36**.

Limpscomb, T.P., Kennedy, S., Moffett, D. and Ford, B.K. (1994) Morbilliviral disease in Atlantic bottlenose dolphins (*Tursiops truncatus*) from 1987-1988 epizootic. *Journal of Wildlife Diseases*, **30**, 285-292.

Limpscomb, T.P., Schulman, F.Y., Moffett, D. and Kennedy, S. (1994) Morbilliviral disease in Atlantic bottlenose dolphins (*Tursiops truncatus*) from the 1987-1988 epizootic. *Journal of Wildlife Diseases*, **30**, 567-571.

Limpscomb, T.P. (2000) Infectious diseases. In Davis, C.L. (ed.), *Pathology of Marine Mammals*. AFIP, Washington, D.C.

Lively, C.M. and Dybdahl, M.F. (2000) Parasite adaptation to locally common genotypes. *Nature*, **335**, 268-271.

Lochmiller, R.L. and Deerenberg, C. (2000) Trade-offs in evolutionary immunology: just what is the cost of immunity? *Oikos*, **88**, 87-98.

Loehle, C. (1995) Social barriers to pathogen transmission in wild animal population. *Ecology*, **76**, 326-335.

Lohm, J., Grahn, M., Langefors, A., Andersen, O., Storset, A. and Von Schantz, T. (2002) Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. *Proceedings of the Royal Society of London Biological Sciences*, **269**, 2029-2033.

Long, J.C. (1986) The allelic correlation structure of Gianj- and Kalam- speaking people. I. The estimation and interpretation of Wright's F-statistics. *Genetics*, **112**, 629-642.

Longeri, M., Zanotti, M. and Damiani, G. (2002) Recombinant DRB sequences produced by mismatch repair of heteroduplexes during cloning in *Escherichia coli*. *European Journal of Immunogenetics*, **29**, 517-523.

Lopez de Castro, J.A., Strominger, J.L., Strong, D.M. and Orr, H.T. (1982) Structure of crossreactive human histocompatibility antigens HLA-A28 and HLA-A2: possible implications for the generation of HLA polymorphism. *Proceedings of*

the National Academy of Sciences, **79**, 3813-3817.

Lydyard, P., Whelan, A. and Fanger, M.W. (2004) *Immunology*. BIOS Scientific Publishers, London.

Mamaev, L.V., Denikina, N.N., Belikov, S.I., Volchkov, V.E., Visser, I.K.G., Fleming, M., Kai, C., Harder, T.C., Liess, B., Osterhaus, A.D.M.E. and Barrett, T. (1995) Characterisation of morbilliviruses isolated from Lake Baikal seals. *Veterinary Microbiology*, **44**, 251-259.

Mamaev, L.V., Visser, K.G., Belikov, S.I., Denikina, N.N., Harder, T., Goatley, L., Rima, B., Edginton, B., Osterhaus, A.D.M.E. and Barrett, T. (1996) Canine distemper virus in Lake Baikal seals (*Phoca sibirica*). *Veterinary Record*, **138**, 437-439.

Markow, T., Hedrick, P.W., Zuerlein, K., Danilovs, J., Martin, J., Vyvial, T. and Armstrong, C. (1993) HLA polymorphism in the Havasupai: evidence for balancing selection. *American Journal of Human Genetics*, **53**, 943-952.

Martin, A.R. and Reeves, R.R. (2002) Diversity and zoogeography. In Hoelzel, A.R. (ed.), *Marine mammal biology; an evolutionary approach*. Blackwell Science Ltd, Oxford.

Mayer, W.E., Jonker, D., Klein, D., Ivanyi, P., Van Seventer, G. and Klein, J. (1988) Nucleotide sequence of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. *EMBO J*, **7**, 2765-2774.

McClelland, E.E., Penn, D.J. and Potts, W.K. (2003) Major histocompatibility complex heterozygote superiority during coinfection. *Infestion and Immunity*, **71**, 2079-2086.

McDade, T.W. and Worthman, C.M. (1999) Evolutionary process and the ecology of human immune function. *American Journal of Human Biology*, **11**, 705-717.

McDevitt, H. (1995) Evolution of MHC Class II Allelic Diversity. *Immunological Reviews*, **143**, 114-122.

Mead, J.G. (2002) Beaked Whales. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Meyer, D. and Thomson, G. (2001) How selection shapes variation of the human major histocompatibility complex:a review. *Annual Human Genetics*, **65**, 1-26.

Migaki, G., Sawa, T.R. and Dubey, J.P. (1990) Fatal Disseminated Toxoplasmosis in a Spinner Dolphin. *Veterinary Pathology*, **27**, 463-464.

Mikaelian, I., Labelle, P., Dore, M. and Martineau, D. (1999) Metatastic mammary adenocarcinomas in two beluga whales (*Delphinapterus leucas*) from the St Lawrence estuary, Canada. *Veterinary Record*, **145**, 738-739.

Mikaelian, I., Boisclair, J., Dubey, J.P., Kennedy, S., Martineau, D. (2000) Toxoplasmosis in Beluga Whales (*Delphinapterus leucas*) from the St Lawrence estuary: Two Case Reports and a Serological Survey. *Journal of Comparative Pathology*, **122**, 73-76.

Mikko, S. and Andersson, L. (1995) Low major histocompatibility complex class II diversity in European and North American moose. *Proceedings of the National Academy of Sciences*, **92**, 4259-4263.

Minh Binh, T., Watanabe, M., H.Nakata, tanabe, S. and Jefferson, T.A. (1999) Contamination by persistent organochlorides in small cetaceans from Hong Kong coastal waters. *Marine Pollution Bulletin*, **39**, 383-392.

Miyamoto, M.M. and Fitch, W.M. (1996) Constraints on protein evolution and the age of the eubacteria/ eukaryote split. *Systematic Biology*, **45**, 568-575.

Miyata, T., Miyazawa, S. and Yasunaga, T. (1979) Two types of amino acid substitutions in protein evolution. *Journal of Molecular Evolution*, **12**, 219-236.

Miyata, T., Yasunaga, T. and Nishida, T. (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *Journal of Molecular Evolution*, **16**, 23-36.

Muller, G., Siebert, U., Wunschmann, A., Artelt, A. and Baumgartner, W. (2000) Immunohistological and serological investigation of morbillivirus infection in harbour porpoises (*Phocoena phocoena*) from the German Baltic and North Sea. *Veterinary Microbiology*, **75**, 17-25.

Murray, J.D. (1993) *Mathematical biology*. Springer-Verlag, Heilderg.

Murray, B.W., Malik, S. and White, B.N. (1995) Sequence variation at the Major Histocompatibility Complex Locus *DQb* in Beluga Whales (*Delphinapterus leucas*). *Molecular Biology and Evolution*, **12**, 582-593.

Murray, B.W., Michaud, R. and White, B.N. (1999) Allelic and Haplotype variation of MHC cclass II *DRB1* and *DQB* loci in St Lawrence beluga (*Delphinapterus leucas*). *Molecular Ecology*, **8**, 1127-1139.

Muruyama, T. and Nei, M. (1981) Genetic variability maintained by mutations and overdominant selection in finite populations. *Genetics*, **98**, 441-459.

Mylius, S.D. and Metz, J.A.J. (2004) When does evolution optimize? on the relationship between evolutionary stability, optimization and density dependence. In Dieckmann, U. and Metz, J.A.J. (eds.), *Elements of adaptive dynamics*. Cambridge University Press, Cambridge.

Nasir, L., Ndiaye, M., Seely, C. and Stear, M.J. (1997) Sequence polymorphism in the bovine major histocompatibility complex *DQB* loci. *Animal Genetics*, **28**, 441-445.

Natoli, A., Peddemors, V.M. and Hoelzel, A.R. (2004) Population structure and speciation in the genus *Tursiops* based on microsatellite and mitochondrial DNA analyses. *Journal of Evolutionary Biology*, **17**, 363-375.

Nei, M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, **70**, 3321-3323.

Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418-426.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.

Nei, M. (1991) Relative efficiencies of different tree making methods for molecular data. In Miyamoto, M.M. and Cracraft, J.L. (eds.), *Recent advances in phylogenetic studies of DNA sequences*. Oxford University Press, Oxford.

Nei, M., Ping, X. and Galina, G. (2000) Estimation of divergence times from mulitprotein sequences for a few mammalian species and several distantly related organisms. *Proceedings of the National Academy of Sciences*, **98**, 2497-2502.

Nepom, G.T. and Erlich, H.A. (1991) MHC class II molecules and autoimmunity. *Annual Review in Immunology*, **9**, 493-525.

Nielsen, R. and Yang, Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, **148**, 929-936.

Nikaido, M., Rooney, A.P. and Okada, N. (1999) Phylogenetic relationships among interspersed elements; hippopotamuses are the closest living relatives to whales. *Proceedings of the National Academy of Sciences*, **96**, 10261-10266.

Nonaka, M., Namikawa, C., Kato, Y., Sasaki, M. and Salter-Cid, L. (1997) Major histocompatibility complex gene mapping in the amphibian *Xenopus* implies a primordial organization. *Proceedings of the National Academy of Sciences*, **94**, 5789-5791.

Nunn, C.L., Gittleman, J.L. and Antonovics, J. (2000) Promiscuity and the primate immune system. *Science*, **290**, 1168-1170.

Ober, C., Weitkamp, L.R., Cox, N., Dytch, H., Kostyu, D. and Elias, S. (1997) HLA and mate choice in humans. *American Journal of Human Genetics*, **61**, 497-504.

O'Brien, S.J., Roelke, M.E., Marker, L., Newman, A., Winkler, C.A., Meltzer, D., Colly, L., Evermann, J.F., Bush, M. and Wildt, D.E. (1985) Genetic basis for species vulnerability in the Cheetah. *Science*, **227**, 1428-1434.

O'Brien, S.J. (1994) A role for molecular genetics in biological conservation. *Proceedings of the National Academy of Sciences*, **91**, 5748-5755.

O'Corry-Crowe, G.M. (2002) Beluga Whale. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

O'Hara, T.M., House, C., House, J.A., Suydam, R.S. and George, J.C. (1998) Viral serologic survey of Bowhead whales in Alaska. *Journal of Wildlife Diseases*, **34**, 39-46.

O'hUigin, C. (1995) Quantifying the Degree of Convergence in Primate Mhc-DRB Genes. *Immunological Reviews*, **143**, 123-140.

Olson, R.R., De Magistris, M.T., Di Tommaso, A. and Karr, R.W. (1992) Mutations in the third, but not the first or second, hypervariable regions of DR(1*0101) eliminate DR1-restructed recognition of a *pertussis* toxin peptide. *Journal of Immunology*, **148**, 2703-2708.

Olson, R.R., Reuter, J.J., McNicholl, J., Alber, C., Klohe, E., Callahan, K., Siliciano, R.F. and Karr, R.W. (1994) Acidic residues in the DR□chain third hypervariable region are required for stimulation of a DR(□, □1*0402)-restricted T-cell clone. *Human Immunology*, **41**, 193-200.

Olson, P.A. and Reilly, S.B. (2002) Pilot whales. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Orvell, C. and Sheshberadaran, H. (1991) Phocine distemper virus is phylogenetically related to canine distemper virus. *Veterinary Record*, **129**, 267-269.

Osterhaus, A.D.M.E., de Swart, R.L., Vos, H.W., Ross, P.S., Kenter, M.J.H. and Barrett, T. (1995) Morbillivirus infections of aquatic mammals: newly identified members of the genus. *Veterinary Microbiology*, **44**, 219-227.

Otting, N., De Groot, N.G., CDoxiadis, G.G.M. and Bontrop, R.E. (2002) Extensive *Mhc-DQB* variation in humans and non-human primate species. *Immunogenetics*, **54**, 230-239.

Ou, D., Mitchell, L.A., Domeier, M.E., Tsang, A.O.W., Decarie, D., Nepom, G.T., Lacroix, M. and Zrein, M. (1996) Characterization of HLA restrictive elements of rubella virus-specific cytotoxic T-cell clone: influence of HLA-DR4b chain residue 74 polymorphism on Ag peptide-T cell interaction. *International Immunology*, **8**, 1577-1586.

Ou, D., Mitchell, L.A. and Tingle, A.J. (1997) HLA-DR restrictive supertype dominate promiscuous T cell recognition: association of multiple HLA-DR molecules with susceptibility to autoimmune diseases. *Journal of Rheumatology*, **24**, 253-261.

Ou, D., Mitchell, L.A. and Tingle, A.J. (1998) A new categorization of HLA DR alleles on a functional basis. *Human Immunology*, **59**, 665-676.

Page, R.D.M. and Holmes, E.C. (1998) *Molecular Evolution; a phylogenetic approach.* Blackwell Science Ltd.

Palmer, C.J., Schroeder, J.P., Fujioka, R.J. and Douglas, J.T. (1991) *Staphylococcus aureus* infection in newly captured Pacific bottlenose dolphins (*Tursiops truncatus gilli*). *Journal of Zoo and Wildlife Medicine*, **22**, 330-338.

Pamilo, P. (1984) Genotypic correlation and regression in social groups: multiple alleles, multiple loci and subdivided populations. *Genetics*, **107**, 307-320.

Pamilo, P. and Bianchi, N.O. (1993) Evolution of Zfx and Zfy genes: rates and interdependence between genes. *Molecular Biology and Evolution*, **10**, 271-281.

Parham, P., Lomen, C.E., Lawlor, D.A., Ways, J.P., Holmes, N., Coppin, H.L., Salter, R.D., Wan, A.M. and Ennis, P.D. (1988) Nature of polymorphism in HLA-A, -B and -C molecules. *Proceedings of the National Academy of Sciences*, **85**, 4005-4009.

Paterson, S., Wilson, K. and Pembertob, J.M. (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). *Proceedings of the National Academy of Sciences*, **95**, 3714-3719.

Paul, W.E. (ed.). (1993) *Fundamental Immunology*. Raven Press, New York.

Penn, D.J., Damjanovich, K. and Potts, W.K. (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences*, **99**, 11260-11264.

Perrin, W.F. and Brownell, R.L. (2002) Minke whales. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Pillet, S.V., Lesage, M., Hammill, D., Cyr, G., Bouquegneau, J.-M. and Fournier, M. (2000) *In Vitro* exposure of seal peripheral blood leukocytes to different metals reveal a sex-dependent effect of zinc on phagocytic activity. *Marine Polution Bulletin*, **40**, 921-927.

Potts, W.K., Manning, C.J. and Wakeland, E.K. (1991) Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature*, **352**, 619-621.

Potts, W.K. and Slev, P.R. (1995) Pathogen-Based Models Favoring MHC Genetic Diversity. *Immunological Reviews*, **143**, 181-197.

Pupko, T., Pe'er, I., Shamir, R. and Graur, D. (2000) A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Molecular Biology and Evolution*, **17**, 890-896.

Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258-275.

Raymont, M. and Rousset, F. (1995) GENEPOP (version 2.1): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.

Read, A. (2002) Porpoises. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Read, J.M. and Keeling, M.J. (2003) Disease evolution on networks; the role of contact structure. *Proceedings of the Royal Society of London Biological Sciences*, **270**, 699-708.

Read, A. (2004) Annual cetacean bycatch. In IWC (ed.), *12th Annual meeting*, Berlin.

Reeves, R.R. (2002) Conservation efforts. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Richardson, D.S. and Westerdahl, H. (2003) MHC diversity in two *Acrocephalus* species: the outbred Great reed warbler and the inbred Seychelles warbler. *Molecular Ecology*, **12**, 3523-3529.

Rima, B.K., Wishaupt, R.G.A., Welsh, M.J. and Earle, J.A.P. (1995) The evolution of morbilliviruses: a comparison of nucleocapsid gene sequences including a porpoise morbillivirus. *Veterinary Microbiology*, **44**, 127-134.

Rohl, A. (1997) Network: a program package for phylogenetic networks. Universitat Hamburg, Hamburg.

Roitt, I., Brostoff, J. and Male, D. (2001) *Immunology*. Mosby International Ltd., London.

Romano, T.A., Ridgay, S.H. and Quaranta, V. (1992) MHC class II molecules and immunoglobins on peripheral blood lymphocytes of the bottlenose dolphin, *Tursiops truncatus*. *Journal of Experimental Zoology*, **263**, 96-104.

Ross, P.S., Visser, K.G., Broeders, H.W.J., van de Bilt, M.W.G., Bowen, W.D. and Osterhaus, A.D.M.E. (1992) Antibodies to phocine distemper virus in Canadian seals. *Veterinary Record*, **130**, 514-516.

Ross, P.S., Jahans, K.l., MacMillan, A.P., Reid, R.J., Thompson, P.M., Foster, G. (1996) *Brucella* species infection in North Sea seal and cetacean populations. *Veterinary Record*, **138**, 647-648.

Ross, P.S., De Swart, R.L., Van Loveren, H., Osterhaus, A.D.M.E. and Vos, J.G. (1996) the immunotoxicity of environmental contaminants to marine wildlife: a review. *Annual Review of Fish Diseases*, **6**, 151-165.

Ross, P.S., Ellis, G., Ford, J.K.B. and Barrett-Lennard, L. (2002) Toxic chemical

pollution and pacific killer whlaes (*Orcinus orca*). *Fourth International Orca Symposium and Workshops*, France, pp. 126-130.

Rousset, F. (1997) Genetic differentiation of gene flow from *F*-Statistics under isolation by distance. *Genetics*, **145**, 1219-1228.

Russel, C.A., Smith, D.L., Waller, L.A., Childs, J.E. and Real, L.A. (2003) A priori prediction of disease invasion dynamics in a novel environment. *Proceedings of the Royal Society of London Biological Sciences*, **271**, 21-25.

Saitou, N. and Nei, M. (1987) The neighbour-joining method: a new method for reconstracting phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406-425.

Sambrook, J., Fritsch, E.F. and T., M. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press.

Sanchez-Mazas, A., Djoulah, S., Busson, M., Le Monneir de Gouville, I., Poirier, J., Dehay, C., Charron, D., Excoffier, L., Schneider, S., Langaney, A., Dausset, J. and Hors, J. (2000) A linkage disequilibrium map of the MHC region based on the analysis of 14 loci hapltotypes in 50 French families. *European Journal of Immunogenetics*, **8**, 33-41.

Sanjeevi, C.B., Shtaurere, A., Samal, K.C. and Tripath, B.B. (1999) Association of HLA class II alleles with different subregions of diabetes mellitus in Eastern India identify different associations with IDDM and malnutrition related diabetes. *Tissue Antigens*, **54**, 83-87.

Sant, A.J. (1994) Endogenous antigen presentation by MHC class II molecules. *Immunological Research*, **13**, 253-267.

Satta, Y., O'Huigin, C., Takahata, N. and Klein, J. (1994) Intensity of natural selection at the major histocompatibility complex loci. *Proceedings of the National Academy of Sciences*, **91**, 7184-7188.

Schmid-Hempel, P. (2003) Variation in immune defence as a question of evolutionary ecology. *Proceedings of the Royal Society of London Biological Sciences*, **270**, 357-366.

Schneider, S., Vincek, V., Tichy, H., Figuera, F. and Klein, J. (1991) MHC Class II genes of a marsupial, the red-necked Wallaby (*Macropus rufogriseus*): identification of new gene families. *Molecular Biology and Evolution*, **8**, 753-766.

Schneider, S., Roessli, D. and Excoffier, L. (2000) ARLEQUIN: a software for population genetics data analysis. University of Geneva, Geneva.

Scott, M.D. and Wells, R.S. (2002) Bottlenose dolphins. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Seddon, J.M. and Baverstock, P.R. (1999) Variation on islands: major

histocompatibility complex (*Mhc*) polymorphism in populations of the Australian bush rat. *Molecular Ecology*, **8**, 2071-2079.

Seddon, J.M. and Ellegren, H. (2002) MHC class II genes in European wolves: a comparison with dogs. *Immunogenetics*, **54**, 490-500.

Sempel, S., Cowlishaw, G. and Bennet, P.M. (2001) Immune system evolution among anthropoid primates: parasites, injuries and predators. *Proceedings of the Royal Society of London Biological Sciences*, **269**, 1031-1037.

Sette, A., Buss, S., Colon, S., Smith, J.A., Miles, C. and Grey, H.M. (1987) Structural characteristics of an antigen required for its interaction with Ia and recognition by T cells. *Nature*, **328**, 395-399.

Shadan, F.F. and Villarreal, L.P. (1993) Coevolution of persistently infecting small DNA viruses and their hosts linked to host-interactive regulatory domains. *Proceedings of the National Academy of Sciences*, **90**, 4117-4121.

Shcmid-Hempel, P. (2003) variation in immune defence as a question of evolutionary ecology. *Proceedings of the National Academy of Sciences*, **270**, 357-366.

She, J.X., Boehme, S.A., Wang, T.W., Bonhomme, F. and Wakeland, E.K. (1991) Amplification of major histocompatibility complex class II gene diversity by intraexonic recombination. *Proceedings of the National Academy of Sciences*, **88**, 453-457.

Slade, R.W. (1992) Limited MHC polymorphism in the southern elephant seal: implications for MHC evolution and marine mammal population biology. *Proceedings of the Royal Society of London Biological Sciences*, **249**, 163-171.

Slade, R.W. and McCallum, H.I. (1992) Overdominant *vs.* Frequency-Dependent selection at MHC Loci. *Genetics*, **132**, 861-862.

Smith, N.G.C. (2003) Are radical and conservative substitution rates useful statistics in molecular evolution? *Journal of Molecular Evolution*, **57**, 467-478.

Soeda, E., Maruyama, T., Arrand, J.R. and Griffin, B.E. (1980) Host-dependent evolution of three papova viruses. *Nature*, **285**, 165-167.

Sokal, R.R. and Rohl, F.J. (1995) *Biometry*. Freeman and Co., New York.

Sommer, S. (2003) Effects of habitat fragmentation and changes of dispersal behavioru after a recent population decline on the genetic variability of noncoding and coding DNA of a monogamous Malagasy rodent. *Molecular Ecology*, **12**, 2845-2851.

Sorci, G., Morand, S. and Hugot, J.P. (1997) Host-parasite coevolution: comparative evidence for covariation of life-history traits in primates and oxyrurid parasites. *Proceedings of the Royal Society of London Biological Sciences*, **264**, 285-289.

Stern, J.L., H., B.J., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L. and Wiley, D.C. (1994) Crystal structure of the human class II MHc protein HLA-DRB1 complexed with an influenza virus peptide. *Nature*, **368**, 215-221.

Stet, R.J.M. and Egberts, E. (1991) The histocompatibility system in teleostean fishes: from multiple loci to a major histocompatibility complex. *Fish and Shellfish Immunology*, **1**, 1-10.

Stuart, M.D. and Strier, K.B. (1995) Primates and parasites: a case for a multidisciplinary approach. *International Journal of Primatology*, **16**, 577-593.

Stuart, A., Ord, J.K. and Arnold, S. (1999) *Kendall's advanced theory of statistics*. Arnold Publishers.

Swarbick, P.A. and Crawford, A.M. (1997) The red deer (*Cervus elaphus*) contains two expressed major histocompatibility complex class II *DQB* genes. *Animal Genetics*, **28**, 49-51.

Swofford, D.L. (1991) PAUP: phylogenetic analysis using parsimony. Illinois Natural History Survey, Champain, Illinois.

Taddel, F., Scarelli, V., Frenzilli, G. and Nigro, M. (2001) genotoxic hazard of pollutants in cetaceans: DNA damage and repair evaluated in the bottlenose dolphin (*Tursiops truncatus*) by the Connet assay. *Marine Pollution Bulletin*, **42**, 324-328.

Takahata, N. and Nei, M. (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, **124**, 967-978.

Takahata, N., Satta, Y. and Klein, J. (1992) Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics*, **130**, 925-938.

Tamura, K. and Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512-526.

Taubenberger, J.K., Tsai, M.M., Atkin, J.T., Fanning, T.G., Krafft, A.E., Moeller, R.B., Kodsi, S.E., Mense, M.G. and Lipscomb, T.P. (2000) Molecular Genetic Evidence of a Novel Morbillivirus in a Long-Finned Pilot Whale (*Globicephalus melas*). *Emerging Infectious Diseases*, **6**, 42-45.

Taylor, B.L. (2002) Conservation biology. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Thewissen, J.G.M., Hussain, S.T. and Arif, M. (1994) Fossil evidence for the origin of aquatic locomotion in *Archaeocete* whales. *Science*, **263**, 210-212.

Thewissen, J.G.M., Madar, S.I. and Hussain, S.T. (1996) *Ambulocetus natans*, an Eocene cetacean (Mammalia) from Pakistan. *Courier Forschungs-Institut*

Senckenberg, **190**.

Thewissen, J.G.M. and Hussein, S.T. (1998) Systematic review of the Packicetidae, Early and Middle Eocene Cetacea from Pakistan and India. *Bulletin of the Carnegie Museum of Natural History*, **34**, 220-238.

Thewissen, J.G.M. (1998) Cetacean origins: evolutionary turnmoil during the invasion of the oceans. In Thewissen, J.G.M. (ed.), *The emergence of whales: evolutionary patterns in the origin of Cetacea*. Plenum press, New York.

Townsend, A., Ohlen, C., Bastin, J., Ljunggren, H.G., Foster, L. and Karre, K. (1989) Association of Class I major histocompatibility heavy and loght chains induced by viral peptides. *Nature*, **340**, 443-451.

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T., Hayes, E., Funkhouser, R., Fugate, M., Theiler, J., Hsu, Y.S., Kunstman, K., Wu, S., Phair, J., Erlich, H.A. and Wolinsky, S. (2003) Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, **9**, 928-935.

Trowsdale, J., Grove, V. and Arnason, A. (1989) MHC polymorphism in whales. *Immunogenetics*, **29**, 19-24.

Trowsdale, J. (1995) 'Both man & bird & beast': comparative organization of MHC genes. *Immunogenetics*, **41**, 1-17.

Trowsdale, J. and Parham, P. (2004) Defence strategies and immunity-related genes. *European Journal of Immunology*, **34**, 7-17.

Trudgett, A., Lyons, C., Welsh, M.J., Duffy, N., McCullough, S.J. and McNeilly, F. (1991) Analysis of a seal and a porpoise morbillivirus using monoclonal antibodies. *Veterinary Record*, **128**, 61.

Tsuji, K., Aizawa, M. and Sasazuki, T. (1992) Proceedings of the 11th international histocompatibility workshop and conference. Oxford University Press.

Tufto, J., Engen, S. and Hindar, K. (1996) Inferring patterns of migration from gene frequencies under equilibrium conditions. *Genetics*, **144**, 1911-1921.

Tyack, P.L. and Whitehead, H. (1983) Male competition in large groups of wintering Humpback whale. *Behavior*, **83**, 1-23.

Udina, I.G., Sokolova, S.S., Sipko, T.P. and Sulimova, G.E. (1994) Comparative characteristics of DNA polymorphism for major histocompatibility complex loci DQB and DRB in representatives of the *Bovidae* family. *Genetika*, **30**, 356-360.

Urban, R.G. and Chicz, R.M. (eds.). (1996) *MHC molecules: expression, assembly and function*. R.G Landes Company, Chapman and Hall.

Valdes, A.M., McWeeney, S.K. and Thomson, G. (1999) Locus and population specific evolution in HLA class II genes. *Annual Human Genetics*, **63**, 27-43.

Valsescchi, E., Hale, P., Corkeron, P. and Amos, W. (2002) Social structure in migrating humpback whales (*Megaptera novaeangliae*). *Molecular Ecology*, **11**, 507-518.

Van Boven, M. and Weissing, F.J. (2004) The Evolutionary Economics of Immunity. *The American Naturalist*, **163**, 277-294.

Van Bressem, M.F., Visser, I.K.G., Van De Bilt, M.W.G., Teppema, J.S., Raga, J.A. and Osterhaus, A.D.M.E. (1991) Morbillivirus infection in Mediterranean striped dolphins. *Veterinary Record*, **129**, 471-472.

Van Bressem, M.F., De Meurichy, J., Chappuis, G., Spehner, D., Kieny, M.P., Pastoret, P.P. (1991) Attempt to vaccinate orally harbour seals against phocid distemper. *Veterinary Record*, **129**, 362.

Van Bressem, M.F., Van Waerebeek, K., Reyes, J.C., Dekegel, D. and Pastoret, P. (1993) Evidence of Poxvirus in Dusky Dolphin (*Lagenorhynchus obscurus*) and Burmeister's porpoise (*Phocoena spinipinnis*) from Coastal Peru. *Journal of Wildlife Diseases*, **29**, 109-113.

Van Bressem, M.F., Van Waerebeek, K., Flemming, M. and Barrett, T. (1998) serological evidence of morbillivirus infection in small cetaceans from the Southeast Pacific. *Veterinary Microbiology*, **59**, 89-98.

Van Bressem, M.F., Van Waerebeek, K. and Raga, A.J. (1999) A review of virus infections of cetaceans and the potential impact of morbilliviruses , poxviruses and pappilomaviruses on host population dynamics. *Diseases of Aquatic Organisms*, **38**, 53-65.

Van Bressem, M. (1999) Further insight on the epidemiology of cetacean morbillivirus in the Northeastern Atlantic. *Marine Mammal Science*, **14**, 605-613.

Van Bressem, M.F., Kastelein, R.A., Flamant, P. and Orth, G. (1999) Cutaneous papillomavirus infection in a harbour porpoise (*Phocoena phocoena*) from the North Sea. *Veterinary Record*, **144**, 592-593.

Van Bressem, M.F., Van Waerebeek, K., Jepson, P.D., Raga, A.J., Duignan, P.J., Nielsen, O., Di Beneditto, A.P., Siciliano, S., Ramos, R., Kant, W., Peddemors, V.M., Kinoshita, R., Ross, P.S., Lopez-Fernandez, A., Evans, K., Crespo, E. and Barrett, T. (2001) An insight into the epidemiology of dolphin morbillivirus worldwide. *Veterinary Microbiology*, **81**, 287-304.

Van De Bilt, M.W.G., Martina, B.E.E., Vedder, E.J., Androukaki, E., Kotomatas, S., Komnenou, A., Sidi, B.A., Jiddou, A.B., Barham, M.E.O., Niesters, H.G.M. and Osterhaus, A.D.M.E. (2000) Identification of morbilliviruses of probable cetacean origin in carcases of Mediterranean monk seals (*Monachus monachus*). *Veterinary Record*, **146**, 691-694.

Van Den Bussche, R.A., Ross, T.G. and Hoofer, S.R. (2002) Genetic variation at a

major histocompatibility locus within and among populations of white-tailed deer (*Odocoileus virginianus*). *Journal of Mammalogy*, **83**, 31-39.

Van Der Walt, J.M., Nel, L.H. and Hoelzel, A.R. (2001) Characterization of major histocompatibility complex DRB diversity in the endemic South African antelope *Damaliscus pygargus*: a comparison in two subspecies with different demographic histories. *Molecular Ecology*, **10**, 1679-1688.

Van Pihlen, I.A., Amos, B. and Burke, T. (1995) Patterns of genetic variability at individual minisatellite loci in minke whale *Balaenoptera acutorostrata* populations from three different oceans. *Molecular Biology and Evolution*, **12**, 459-472.

Verspoor, E. and Jordan, W.C. (1989) Genetic variation at the Me-2 locus in the Atlantic salmon within and between rivers: evidence for its selective maintenance. *Journal of Fish Biology*, **35**, 205-213.

Villegas-Castagnasso, E.E., Diaz, S., Ciovambattista, G., Dulout, F.N. and Peral-Garcia, P. (2003) Analysis of ELA-DQB exon-2 polymorphism in Argentine Creole horses by PCR-RFLP and PCR-SSCP. *Journal of Veterinary Medicine*, **50**, 280-285.

Visser, I.K.G., Van Bressem, M.F., De Swart, R.L., van de Bilt, M.W.G., Vos, H.W. and Van De Heijden, R.W.J. (1993) Characterization of morbilliviruses isolated from dolphins and porpoises in Europe. *Journal of General Virology*, **74**, 631-641.

Vogelstein, B. and Gillespie, D. (1979) Preparative and analytical purification of DNA from agarose. *Proceedings of the National Academy of Sciences*, **76**, 615-619.

Vossen, M.T.M. (2002) Viral immune evasion: a masterpiece of evolution. *Immunogenetics*, **54**, 527-542.

Wagner, J.L., Hayes-Lattin, B., Works, J.D. and Storb, R. (1998) Molecular analysis and polymorphism of the DLA-*DQB* genes. *Tissue Antigens*, **52**, 242-250.

Wang, J.Y., Chou, L.S. and White, B.N. (1999) Mitochondrial DNA analysis of sympatric morphotypes of bottlenose dolphins (genus: *Tursiops*) in Chinese waters. *Molecular Ecology*, **8**, 1603-1612.

Watterson, G.A. (1978) The homozygosity test of neutrality. *Genetics*, **88**, 405-417.

Webber, D.S., Stewart, B.S., Schienman, J. and Lehman, N. (2004) Major histocompatibility complex variation at three class II loci in the northern elephant seal. *Molecular Ecology*, **13**, 711-718.

Wecherpfennig, K.W. and Strominger, J.L. (1995) Selective binding of self peptides to disease-associated major histocompatibility complex (MHC) molecules: mechanisms for MHC-linked susceptibility to human autoimmune diseases. *Journal of Immunology*, **181**, 1597-1602.

Wecherpfennig, K.W., Yu, B., Bhol, K., Monos, D.S., Argyris, E., Karr, R.W., Ahmed, A.R. and Strominger, J.L. (1995) Structural basis for major histocompatibility complex(MHC)-linked susceptibility to autoimmunity: charged residues of a single MHC binding pocket confer selective presentation of self-peptides in pemphigus vulgaris. *Proceedings of the National Academy of Sciences*, **92**, 11935-11939.

Wedekind, C., Seebeck, T., Bettens, F. and Paepke, A.J. (1995) MHC -dependent mate preferneces in humans. *Proceedings of the Royal Society of London Biological Sciences*, **260**.

Wedekind, C. and Furi, S. (1997) Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity. *Proceedings of the Royal Society of London Biological Sciences*, **264**, 1471-1479.

Wegner, K.M., Reusch, T.B.H. and Kalbe, M. (2003) Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology*, **16**, 224-232.

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

Weir, B.S. and Hill, W.G. (2002) Estimating F-statistics. *Annual Reviews in Genetics*, **36**, 721-750.

Wenink, P.W., Groen, A.F., Roelke-Parker, M.E. and Prins, H.H.T. (1998) African buffalo maintain high genetic diversity in the major histocompatibility complex in spite of historically known population bottlenecks. *Molecular Ecology*, **7**, 1315-1322.

Westerdahl, H., Hansson, B., Bensch, S. and Hasselquist, D. (2004) Between-year variation of MHC allele frequencies in great reed warblers: selection or drift. *Journal of Evolutionary Biology*, **17**, 485-492.

White, K.A.J., Lewis, M.A. and Murray, J.D. (1996) A model of wolf-pack territory formation and maintenance. *Journal of Theoretical Biology*, **178**, 29-43.

Whitehead, H. and Weilgart, L. (2000) The sperm whale: social females and roving males. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Cetacean Societies*. University of Chicago Press, Chicago.

Whitehead, H. (2002) Sperm whale. In Perrin, W.F., Wursig, B. and Thewissen, J.G.M. (eds.), *Encyclopaedia of Marine Mammals*. Academic Press, San Diego.

Wild, T.F., Naniche, D., Rabourdin-Combe, C., Gerlier, D., Malvoisin, E., Lecouturier, V. and Buckland, R. (1995) Mode of entry of morbilliviruses. *Veterinary Microbiology*, **44**, 267-270.

Williams, T.M. and Worthy, G.A.J. (2002) Anatomy and Physiology: the challenge of

aquatic living. In Hoelzel, A.R. (ed.), *Marine mammal biology; an evolutionary approach*. Blackwell Science Ltd, Oxford.

Wills, C. (1991) Maintenance of Multiallelic Polymorphism at the MHC region. *Immunological Reviews*, **124**, 165-220.

Wilson, M.L., Bretsky, P.M., Cooper, G.H., Egbertson, S.H., Van Kruinigen, H.J. and Cartter, M.L. (1997) Emergence of Racoon rabies in Connecticut, 1991-1994: spatial and temporal characteristics of animal infection and human contact. *American Journal of Tropical Medicine and Hygiene*, **57**, 457-463.

Winchester, R. (1994) The molecular basis of susceptibility to rheumatoid arthritis. *Advances in Immunology*, **56**.

Worebey, M. (2001) A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria, and mitochondria. *Molecular Biology and Evolution*, **18**, 1425-1434.

Wright, S. (1951) The genetic structure of populations. *Annals of Eugenics*, **15**, 323-354.

Yamazaki, K., Beauchamp, G.K., Egorov, I.K., Bard, J., Thomas, L. and Boyse, E.A. (1983) Sensory distinction between H-2b and H-2bm1 mutant mice. *Proceedings of the National Academy of Sciences*, **80**, 5685-5688.

Yang, Z. (1993) Maximum likelihood estimation of phylogeny from DNA sequences when substitute rates differ over sites. *Molecular Biology and Evolution*, **10**, 1396-1401.

Yang, Z. (1994) Maximum likelihood estimation of phylogeny from DNA sequences when substitute rates differ over sites: approximate methods. *Journal of Molecular Evolution*, **39**, 306-314.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Genetics*, **13**, 555-556.

Yang, Z. (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, **15**, 568-573.

Yang, Z. and Nielsen, R. (1998) Synonymous and non-synonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution*, **46**, 409-418.

Yang, Z. and Nielsen, R. (2000) Estimating synonymous and non-synonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, **17**, 32-43.

Yang, Z. and Bielawski, J.P. (2000) Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution*, **15**, 496-502.

Yucki, N. and O'Brien, S.J. (1990) DNA variation of the mammalian histocompatibility

complex reflects genomic diversity and population history. *Proceedings of the National Academy of Sciences*, **87**, 836-840.

Zhang, J. and Madden, T.I. (1997) PowerBlast: a new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Research*, **7**, 649-656.

Zhang, J. and Nei, M. (1997) Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood and distance methods. *Journal of Molecular Evolution*, **44(Suppl. 1)**, S139-S146.

Zhang, J. (2000) Rates of conservative and radical non-synonymous substitutions in mammalian nuclear genes. *Molecular Biology and Evolution*, **50**, 56-68.

Zhang, J., Zhang, Y.-P. and Rosenberg, H.F. (2002) Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genetics*, **30**, 411-415.

Zinkernagel, R.M. and Doherty, P.C. (1974) Restriction of *in vitro* T-cell mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature*, **248**, 701-702.

Zuk, M. and Stoehr, A.M. (2002) Immune defence and host life history. *American Naturalist*, **160**, S9-S22.